DOCUMENT RESUME

ED 164 866 CE 019 010

AUTHOR Donabedian, Avedis
TITLE Needed Research in the Assessment and Monitoring of the Quality of Medical Care. NCHSR Research Report Series.
INSTITUTION National Center for Health Services Research and Development (DHEW/PHS), Rockville, Md.
REPORT NO DHEW-PHS-78-3215
PUB DATE Jul 78
NOTE 46p.
AVAILABLE FROM National Center for Health Services Research, Office of Scientific and Technical Information, 3700 East-West Highway, Room 7-44, Hyattsville, Maryland 20782

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Bibliographies; Evaluation Criteria; Measurement Techniques; *Medical Care Evaluation; Medical Services; Models; Program Evaluation; *Quality Control; Research Methodology; *Research Needs; Standards; State of the Art Reviews

ABSTRACT

In this paper the author assesses the state-of-the-art on quality assessment and monitoring of medical care and makes recommendations for needed research in this area. Following a brief introduction, the content is presented in two sections. The first, providing a frame of reference, covers definitions; quality assessment and program evaluation; relationship of quality and quantity; relationship of quality and cost; strategies of care; structure, process, and outcome; monitoring versus research; and the uses of outcomes. The second section presents a catalog of needed research on assessing and monitoring the quality of medical care. The research areas covered are as follows: basic explorations and studies of what constitutes quality, description of prevalent patterns and strategies of care, the epidemiology of quality, the relationship of structure to process or outcome, development of basic tools for assessment, specification and testing of system-design elements, comparative studies of quality using different approaches, further development of promising current approaches, integrative measures of quality, applications to special areas, consumer perspectives and the consumer's role, quality assessment and monitoring as a social process, and effectiveness and the factors that influence it. An extensive bibliography is attached. (EM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED164866

NCHSR

RESEARCH REPORT SERIES

Needed Research in the Assessment and Monitoring of the Quality of Medical Care

Avedis Donabedian, M.D., M.P.H.
Department of
Medical Care Organization
School of Public Health
The University of Michigan,
Ann Arbor

July 1978

CE 019 010

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service
National Center for Health Services Research

DHEW Publication No. (PHS) 78-3219

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

National Center for Health Services Research

Research Report Series

The *Research Report Series* is published by the National Center for Health Services Research (NCHSR) to provide significant research reports in their entirety. Research Reports are developed by the principal investigators who conduct the research, and are directed to selected users of health services research as part of a continuing NCHSR effort to expedite the dissemination of new knowledge resulting from its project support.

Abstract

The purpose of this paper is to review, evaluate critically, and synthesize the literature on quality assessment and assurance, including the appropriateness of use of service, in order to arrive at a cogent, documented, and authoritative assessment of the state-of-the-art. In addition to addressing quality assessment as a research tool and quality assurance as an administrative tool, an attempt is made to provide an understanding of the epidemiology of quality as a prerequisite to the design of medical care programs and systems. Major components of quality which are discussed include: definitions; quality assessment and program evaluation; relationship of quality and quantity; relationship of quality and cost; strategies of care; structure, process, and outcome; monitoring versus research; and the uses of outcomes. Recommendations for further research in the assessment and monitoring of the quality of medical care are presented.

This *NCHSR Research Report* was written by Avedis Donabedian, M.D., M.P.H., Professor of Medical Care Organization, Department of Medical Care Organization, School of Public Health, The University of Michigan, Ann Arbor. This report, an extended version of a paper prepared for the Department of Medicine and Surgery of the Veterans Administration, is one product of the work currently being supported by the National Center for Health Services Research under grant number HS 02081. The final report from the work supported by NCHSR is not expected to be completed until June 1979. Dr. Donabedian was also supported by The University of Michigan, through sabbatical leave, during part of this study.

Additional copies of this report may be obtained on request from the NCHSR Office of Scientific and Technical Information, 3700 East-West Highway, room 7-44, Hyattsville, MD 20782 (tel: 301/436-8970). Other current NCHSR publications are announced on the last pages of this publication.

Foreword

Prior to 1965, the year in which Medicare became a reality, efforts to improve the quality of medical care were involved mainly with the self-regulation activities of the medical profession, although sporadic research had been conducted in the areas of medical care process and patient outcome as far back as the mid-nineteenth century. Utilization review was emphasized with the passage of Medicare and portended a new awareness of public interest in the quality of care. Dr. Donabedian had the opportunity during the period of 1965-1967 to review and assess the state-of-the-art of quality assessment methodology and responded with several publications which have become classics in the field. Since that time, there has been a period of significant, if not remarkable, growth in the body of knowledge encompassing quality assessment and assurance. A notable development in this area was the implementation of the Experimental Medical Care Review Organization (EMCRO) program in 1971 by the National Center for Health Services Research. This program established the model followed in the development of the Professional Standards Review Organization (PSRO) program and encountered many of the problems subsequently experienced in that program. The EMCRO program addressed problems of criteria development and evaluation, organizational patterns, development of assessment and assurance techniques, impact evaluation, and many of the other emerging issues of the day.

In October, 1972, Public Law 92-603 established formal PSRO review of medical services reimbursed under the Social Security Act, and interest in all facets of quality research received added impetus. The PSRO legislation has the potential for profound impact on the cost and quality of medical services and the form of health services delivery; however, serious concern has been voiced concerning the wisdom of its current mode of implementation, aspects of which have not been rigorously validated.

In the past decade, a considerable body of knowledge has been gathered which requires thoughtful review, evaluation, and synthesis in order to assess the present state-of-the-art and to allow meaningful projections of further research strategies. It is in this framework that Dr. Donabedian suggests that "it is necessary from time to time to pause and take stock of what has been done, so that it may be clearly understood and future effort redirected." The synthesis which he provides is intended to mold together the research, operational, and policy concerns of the health establishment with regard to the quality of medical care at a time when major changes in the financing and organization of health services are on the horizon.

Gerald Rosenthal, Ph.D.
Director
National Center for Health
Services Research

July 1978

Contents

iii FOREWORD
vii INTRODUCTION
1 FRAME OF REFERENCE
Definitions
Quality Assessment and Program Evaluation
2 Quality and Quantity
Quality and Cost
3 Strategies of Care
4 Structure, Process, Outcome
5 Monitoring Versus Research
The Uses of Outcomes

8 PROPOSALS FOR RESEARCH

Basic Explorations and Studies of What Constitutes Quality
Specification and Assessment by Modeling
9 Empirical Testing of Strategies of Care
Increments of Cost and of Quality
10 Studies of the Client-Practitioner Relationship
Views of What Constitutes Quality
Continuity and Coordination as Attributes of Quality

11 Description of Prevalent Patterns and Strategies of Care
Studies of the Elements of Clinical Behavior
Identification of Styles and Strategies
12 Comparisons of Norms with Practice
Homogeneity and Heterogeneity of Performance

The Epidemiology of Quality
13 The Epidemiology of Quality Among Providers
The Epidemiology of Quality Among Clients

14 The Relationship of Structure to Process or Outcome

Development of Basic Tools for Assessment
Specification and Measurement of Outcomes
16 Improvements in the Medical Record
The Criteria and Standards of Quality
17 Monetary Measures of Costs and Benefits

18 Specification and Testing of System-Design Elements
Specifying the Appropriate Cut-Off Points in Standards
Sampling and "Enrichment" Techniques
19 Larger Elements of Design

Comparative Studies of Quality Using Different Approaches

Further Development of Promising Current Approaches
Outcomes as Measures of Quality
Outcomes as Cues and Motivators for Process Assessment

20 The Occurrence of Preventable Adverse Events
The Occurrence of Preventable Progression of Illness or Disability

21 Second Surgical Opinion Programs
Computer-Aided Management
Integrative Measures of Quality

Applications to Special Areas

22 Consumer Perspectives and the Consumer's Role
Representing Client Values in the Definition of Quality
Clients as Sources of Information
Participation in Monitoring
23 Membership in Audit and Utilization Review Committees
Patient Contributions to Care

Quality Assessment and Monitoring as a Social Process

24 Effectiveness and the Factors That Influence It
25 Changes in Physician and Client Behavior
Effect of Technical Design Characteristics
Intra-Institutional "Social Design" Characteristics
26 Supra-Institutional Characteristics and Medical Technology
27 Social-Organizational Characteristics and Medical Technology
Case Studies of Successful and Unsuccessful Programs
Cost-Effectiveness and Cost-Benefit Studies

28 EPILOGUE

29 References
36 Current NCHSR Publications

Introduction

Not too long ago, the quality of medical care was a matter almost exclusively in the professional domain. Any introduction of the subject in a more public setting was to be done, if at all, gingerly, almost apologetically, surrounded by a cloud of caution designed to appease the wrathful and contemptuous professional. How times have changed! "Quality" is now a term perhaps too easily bandied about; and there is little hesitance in proposing that quality can be measured, or that it can be enforced as a matter of public and administrative policy. But this mood of almost belligerent confidence is perhaps premature, for there is much about the concept of quality that is elusive, undefined and unmeasured. Our knowledge of how to go about assuring quality is equally frail.

The academician who seems to be pleading for inaction while proposing further research is a stock figure of ridicule in our gallery of public fools. This is a role I shall try to avoid. My thesis is that while some of us go about doing the best they can with what is known, the effort to examine critically what is being done, and to find new and better ways of doing it, should not be relaxed.

This paper is offered as a modest contribution in this further exploration. It will present a catalog of needed research that is sufficiently organized to avoid being a mere haphazard listing of things. But first, a general framework is needed that explains and justifies the choice of research topics and their organization into a classification. The framework must also indicate what subjects are excluded from consideration. For if reasonably strict limits are not set at the very beginning, the concept of quality has a tendency to expand, so that it embraces every evaluative statement about any aspect of the health care system whatever. Some claim that the concept of quality is enriched when this happens. In my opinionl, this concept can also become attenuated and less useful. Whether I am right or wrong, it is clear that I must circumscribe my subject rather narrowly if I am to complete this paper. For the same reason, the delineation of the frame of reference must also be sketchy, for a thorough specification of the framework is a formidable undertaking in itself. All that is necessary, for now, is that we establish some common ground upon which to build.

A frame of reference

Definitions

In some ways, my definition of quality will be unusually narrow. Quality is taken to signify a judgment on a rather limited segment of the performance of some professional personnel. Primarily, our concern is with the services that are received and consumed personally and directly by individuals, and with those professionals who provide such services. Thus, for instance, environmental services and administrators of personal health services are clearly outside our scope; but the patient-care service of physicians, dentists, pharmacists, nurses, social workers, and other practitioners are clearly included. I believe that it is also reasonable to include the services of professionals, such as pathologists and diagnostic radiologists, who generate the data that are used in direct patient care.

Traditionally, the activities of these several professionals are seen as separate contributions. Hence, one speaks of the quality of physician care, nursing care, social work, and so on. In this paper, I shall follow this tradition by focusing almost exclusively on the services of physicians, but this is not by preference. It is a necessary concession, in part, to the relatively less developed state of quality assessment outside medicine; but, mainly, it is a reflection of my ignorance of much of the work in assessing the quality of performance in the other professions. In fact, an essential component of the concept of quality is the interrelatedness of the contributions of the several professionals in the management of a patient's illness and health. A major item on the research agenda is, therefore, the development of "integrative" measures of quality that take into account this interrelatedness.

Another way in which the concept of quality can be expanded is by accepting a definition of health that is considerably broader than the traditional emphasis on physical-physiological function. I shall include psychological and social well-being as necessary components of health, but only to the extent that responsibility for them can be appropriately undertaken by the physician, under a reasonably liberal definition of his role. It is not useful to define the responsibility of a health care practitioner, for any aspect of health, so that it goes beyond his socially and professionally sanctioned role, or the instrumentalities actually or potentially under his control. 1

Quality assessment and program evaluation

Program evaluation differs from quality assessment by being more inclusive. It deals with activities of health practitioners in addition to direct patient care. It also includes the performance of professionals and nonprofessionals who provide no direct patient care. As a result, there is more to program evaluation than quality assessment; although quality assessment is part of program evaluation, and in many instances its most important part.

In addition to this difference in extent or inclusivity, there are differences that flow from the level of aggregation at which the assessment is performed. Program evaluation deals at the aggregate level with the manner in which a complex formal organization functions, and whether it meets its socially legitimate goals or some specified needs or wants of its clients. In so far as the health care function is concerned, program evaluation places greater emphasis on access to care, on other aspects of resource allocation, on cost and efficiency, and on overall impact on a clientele, a community or a population. By contrast, it has been customary for quality assessment to focus on the appropriateness of care at the level of personal interaction between individual patients and practitioners, as judged by the proper application of science to the needs that the patient and the practitioner have jointly defined. However, as health care becomes more formally organized, and its financing collectivized, there is increasing pressure to introduce into the assessment of quality at the individual level concerns that were previously confined to the collective level of analysis and assessment. This has raised serious ethical problems for the practitioner who is now caught be-

tween the two millstones of responsibility towards the individual patient and the collectivity. It has also blurred formerly clearer distinctions between program evaluation and quality assessment that have now become mostly differences in emphasis and in the details of formulation. Some implications of this blurring and overlap will be clarified in subsequent sections of this paper.

Quality and quantity

The overlap between quantity and quality is one conduit through which program evaluation and quality assessment flow into each other. Obviously, quantitative adequacy is a necessary precondition for quality care. This means that access to care and insufficient use of service are legitimate elements in assessing the quality of client-practitioner transaction as well as of the program as a whole. Insufficient access and use need special and continuing attention because so many programs are obsessively preoccupied with curbing utilization and cutting costs, even though they may pay lip service to the importance of eliminating insufficient care.

Excessive use of service usually means poor quality for a variety of reasons. For one, it implies a lack of skill in the conduct of patient care: an inability to proceed from step to step along the most direct path, selecting the precisely necessary and sufficient procedures to arrive at a diagnosis and to institute treatment. This model of logical and parsimonious progression has traditionally been the hallmark of virtuosity in medical care, and many still hold it as the ideal, although much of current acceptable practice may have departed from it. Note, however, that this is an ideal of minimum redundancy, but not necessarily of minimum cost. It is possible, for example, that using current technology, a strategy of extensive, almost indiscriminate, initial testing will lead to an earlier diagnosis, with lower likelihood of error, at lower average cost, at least in some situations. If so, the criterion of lowest cost could be in conflict with the criterion of least redundancy as elements in the definition of quality. We shall return to this interesting confrontation.

In other ways, the identification of poor quality with excessive use is less problematic. Medical procedures, though intended to be beneficial, are not without risk which, in some cases, is considerable. It is reasonable to assert that the concept of quality includes the criterion that in no case should the benefits expected from any procedure be less than the risks it poses. If so, a problem in assessment is the measurement of risks and benefits in a manner that permits a comparison.

Still another undesirable consequence of excessive use is that by allowing some to have too much, there is less available to others. There is an obvious social misallocation of resources, which is poor quality at the aggregate level. It is comforting to see that, in this instance, the criteria of least redundance at the level of individual patient care and at the collective level call for the same behavior, indicating a harmonious confluence of quality assessment and program evaluation.

One can conclude that any evaluative statement about the quantity of care, beyond a mere description of it, is a judgment about some aspect of quality. For that reason, utilization review and quality assessment are inextricably intertwined, and will be so considered in this paper.

Quality and cost

We have already had intimations that cost is implicated in the concept of quality in a manner that is likely to cause us problems. I am not referring simply to the fact that the empirical relationships between quality and cost are essentially unexplored. I am concerned with the conceptual connections in the definition of quality itself.

The basic considerations that link quality to costs are essentially the same that connect quality to quantity: monetary cost, benefit to health, and risk to health.

There are two ways in which the monetary costs of inputs into care can increase without any increase in the quality of care. First, the elements of service that go into patient care can be provided inefficiently. Thus, hospital care will be costlier than necessary if hospital beds are empty, or the hospital is improperly managed; or if physicians do the work of nurses and the latter the work of aides. Second, the elements of care can be combined and sequenced in a manner that does not realize their full potential to improve health.

The fundamental attribute of these deficiencies in the organization, production and application of care is that there are added costs without either added benefits or added risks. Can this be construed as poor quality of care? I have already argued that at the collective level of analysis the answer could legitimately be in the affirmative because, when resources are scarce, wasteful production and implementation of care reduces the potential benefits of care in the aggregate. The answer could also be in the affirmative at the level of the individual practitioner-client interaction if responsibility for inefficiency falls, at least in part, on the practitioner. To justify this conclusion we can draw on two arguments that I have already used. First, inefficient use of resources suggests lack of skill or judgment in the conduct of care; it is a manifestation of "logical redundancy." Secondly, it is a misallocation of resources, no less at the individual level than in the aggregate. The individual pays more than he ought to for care

either immediately and directly, or in the future as the consequences of program inefficiency eventually work their way back to burden the individual who must, ultimately, foot the bill.

Not everyone will agree that wasteful care is care of poor quality. It is not important to have agreement on this. What is important is that, in evaluations of quality, the element of wastefulness be clearly identified and assessed. Once this is done, the decision could be made either to keep it separate or to merge it into an overall measure of quality.

There is no disagreement that the balance of health benefits and risks is at the core of quality assessment. To do no harm is the most hallowed of precepts in the clinical tradition. No element of care should be used if its risks exceed its expected benefits. Similarly, the combination and sequencing of elements should realize the largest benefits relative to the risks incurred. I shall return to this last stipulation when I deal with the strategies of care.

If quality is measured in terms of actual or expected benefits to health from a specified course of action, the relationship between costs and quality depends on the manner in which the factors I described above enter into care. The elimination or reduction of wastefulness will allow us to have the same quality at lower cost, or higher quality at the same cost. This also holds to the extent that current care includes components that carry unnecessary risk. But the presence of such components adds another, more sinister, aspect to the picture. It is possible for care to actually deteriorate in quality as costs increase above a certain point, if the added services include a large enough component of high-risk, low-benefit items.

Let us assume that the elements of care are produced and applied as efficiently as possible, and that only elements of care are used that have a demonstrable benefit. The relationship between quality and costs, as more and more elements of care are added, is empirically undetermined. One can, however, reasonably assume that the most beneficial elements are likely to be used first and that, as care becomes increasingly elaborate through the addition of more and more elements, the relative gains in benefits and quality become smaller and smaller. In other words, there are diminishing returns and, in the end, very small advantage to be derived from further enrichments in care or additional quantities of it. If we accept this picture, the critical question becomes: At what level of inputs into care should the standard of quality be set? If individual patients paid all the costs and reaped all the benefits the problem would be easily solved by saying that the patient and his physician can jointly decide at what point the added costs are not worth the added benefits. I believe that patients and physicians are capable of making such judgments, using the relatively incomplete information available to them, and when the alternatives are rather clearly demarcated. But in order to make a more precise determination, the data on costs and benefits must be much more accurate, and reducible to a common unit of comparison—for example, dollars.

The solution we have described is clearly in the mainstream of the clinical tradition because it recognizes that patients differ, not only in the strictly "medical" features of their illness, but also in the valuations they place on the costs of care and on its expected benefits. Carried to an extreme, this formulation would support the gut feeling many clinicians have that the quality of care rests on too many individual variations to permit a general standard. What prevents it from going quite that far is the large area of agreement among patients and practitioners on the valuations placed on costs and benefits.

Under the solution proposed above, the quality of care at the collective level would be the sum total of the quality of care determined case by case. But this conclusion could be invalid if the costs were collectivized, for example through health insurance; if the benefits of health care were shared by more than the recipient; and if there were some socially legitimate reason for valuing the health of some, for example children, more than the health of others, for example the aged. To the extent that the collective interest differs from the individual interest, and the health practitioner is made the custodian of both, serious strains are introduced into the patient-practitioner relationship, and the definition of quality poses a moral dilemma.

Strategies of care

Patient care is a planned activity which involves the choice of specific elements of care from a potentially large pool of such elements, and their proper sequencing in order to achieve specified diagnostic and therapeutic objectives. A plan of action, as well as the pattern of actions it generates, may be referred to as a strategy. A very simple example can be assembled from the work of several investigators who have dealt with the choice of strategies for the management of acute pharyngitis in children and adults.

According to Brook, the development of process criteria in cooperation with infectious disease experts resulted in the following recommendations for the treatment of adults with sore throats:

> "If history and physical indicate evidence of a sore throat, a throat culture should be done. If the culture is positive, the patient should make a repeat visit within 24 hours and be given a shot of procaine (short-acting) penicillin; he should then wait 30 minutes in the office so that he can be promptly treated if an anaphylactic reaction occurs. After another 24

hours, the patient should return; if no penicillin allergy has manifested itself in the intervening time, he should be given a shot of benzathine (long-acting) penicillin." [1]

In the example cited by Brook et al. a single strategy is declared, on the authority of experts, to represent "optimal quality of care." Tompkins et al. propose to *test* whether a similar strategy is, in fact, optimal by comparing three hypothetical strategies for the management of sore throat in persons who do not have a history of penicillin allergy or of rheumatic fever. The three alternatives are:

4

"A. the currently recommended strategy, to culture the throat of all patients and to treat with penicillin those whose cultures are positive for group A streptococci;
B. treat all patients with penicillin without taking a throat culture;
C. neither culture not treat any patient." [2]

To complicate matters further, these strategies are tested for epidemic versus endemic occurrence of the disease, for oral versus parenteral penicillin, and for different expected rates of positive throat culture in the population reporting for care.

Forsyth has made the specification of strategies for the management of acute pharyngitis even more complex by introducing as a first step rules that, on the basis of clinical findings, classify cases as (1) most probably caused by group A beta hemolytic streptococcus, (2) most probably not so caused, and (3) questionable as to etiology. With a further subdivision of cases into adults and children six categories of patients are identified: three by etiology and two by age. For each of these it is possible (1) to treat all cases without prior culturing, (2) to culture and then treat, and (3) not to treat at all. As to treatment with penicillin, it can be either oral or parenteral. Thus, there is a matrix of 36 possible combinations, and for each it is possible to test the consequences.[3]

Elaborate as all this seems to be, the experienced clinician will recognize it as an oversimplification of a situation that is itself relatively simple. More complete specification of strategies requires the construction of rather elaborate protocols, algorithms or decision trees. Some reference to this work will be made in a subsequent section. All I want to do do now is to lay down the foundation for asserting that the description and assessment of the elements of care, one at a time, misses the design, the rationale and the implications of the strategy as a whole, including the consequences of taking as well as not taking certain actions. In my opinion, the very essence of quality, that elusive but all-important ingredient that we call clinical judgment, resides in the choice of the most appropriate strategy for the management of any given situation. I also believe that we now have the necessary tools for specifying and testing such strategies, which means that the mysteries of clinical judgment are amenable to yielding their darkest secrets. These tools are decision analysis and the analysis of cost-effectiveness and cost-benefits. But, although these tools are available, the data that are needed for their precise application are generally lacking. We need more accurate information about the occurrence of illness and of clinical and laboratory findings in association with such illness; about the monetary costs and other risks associated with diagnostic procedures, or various sequences of them, in correctly identifying illness when it exists and erroneously missing existing illness; about the risks, burdens and monetary costs of alternative therapies as compared to their contributions to health; about the relative valuations to be placed on various manifestations of health or ill health so that these can be added up into a weighted sum that accurately reflects their total impact; and, where cost-benefit comparisons are to be made, a common unit of measurement is needed that permits a comparison. Even when all the information needed is not available, many insights into clinical judgment can be gained by making use of what is already known, supplemented by the opinions and valuations of those whose practice is under study, and of their patients.

Obviously, a description of the techniques which we have mentioned is beyond the scope of this paper. Lusted has provided a reasonably simple exposition of the clinical applications of decision analysis.[4] A recent paper by McNeil et al. may serve as a quick introduction.[5] The more ambitious reader with an aptitude for quantitative methods is referred to a more rigorous exposition by Raiffa.[6] Klarman has published a brief account of cost-effectiveness and cost-benefit analysis as applied to medical technology, and provided an excellent bibliography.[7] Further discussion of the basic methods referred to above, and a large number of applications to the assessment of strategies in surgical care, are to be found in an excellent book edited by Bunker et al.[8] I shall refer to other work in a subsequent section of this paper; but only the surface can be skimmed, since this is an area currently under intensive investigation. We are, at last, experiencing a major advance in our understanding of key elements in the quality of care. It is a very phlegmatic person indeed who will not be stirred as he first sees, as it were from a mountaintop, this new and enchanting landscape!

Structure, process, outcome

In our frame of reference, when a judgment is made on the quality of care it is taken to be, by definition, a judgment primarily on what professionals do, and how they behave, as they interact

directly with their patients. Hence, it is the process of care that is, ultimately, the object of quality assessment. Quality is defined as the degree of conformance to, or deviation from, normative behavior. In this formulation, both structural attributes and outcomes are indirect means of obtaining information about the normativeness of process.

The rather secondary role assigned to the assessment of structure is not likely to be seriously challenged. Accordingly, in order to make my task manageable, I shall have little to say about structure in this paper, although it will not be entirely excluded. The situation is entirely different with respect to outcomes, since, according to a large body of opinion, including that of many leading experts in quality assessment, it is the outcome of care that is the primary object of concern, and process only a means to the attainment of outcome. It is with diffidence, and with some apology, that a method based on process assessment can venture into this hostile environment; whereas it is a proud badge of honor, assuring almost instant attention and respect, to say that a method is "outcome-oriented." Some have no hesitation to even distort reality, relabeling process elements as outcomes, in order to avoid the obloquy that attaches to process and to bask in the approbation that outcomes confer. Of course, this picture is overdrawn, but not by much!

It is not true that outcomes are a more valid measure of quality than is process, as it is fashionable to say. It is true that process measures are valid indicators of quality only to the extent that they relate to relevant outcomes. But it is equally true that outcomes are valid measures of quality only to the extent that they relate to the antecedent process of care. Fundamentally, validity depends on the strength of the relationship between process and outcome, and on our understanding of that relationship. If that vital link is weak, or in doubt, neither process nor outcome can be used to assess quality; the validity of both is attenuated, and to an equal degree. This means that, in this instance, we do not know what constitutes quality. Further research is needed; the link between process and outcome must be investigated. In such an investigation outcomes are the only reasonable criterion. There can be no disagreement about that, provided the proper outcomes are selected and appropriately measured.

Monitoring versus research

It seems to me that much of the emphasis on the primacy of outcomes in quality assessment arises from a fundamental confusion between research and monitoring. In research, new knowledge is sought about the relationship between outcome and process. In assessment and monitoring, existing knowledge of that relationship is used to obtain information about the behavior of professional personnel or of the larger system.

But, does monitoring have a subsidiary research function? Is monitoring analogous to clinical medicine, where a physician learns how to manage cases by observing the outcomes of care? I think the answer is both "Yes" and "No." It is "Yes" in a restricted sense: if the occurrence of unexpected good or bad outcomes leads to a review of process in the light of currently known relationships between the two; and if, in the event current knowledge does not provide an answer, questions are raised about that knowledge so as to suggest further research. The answer is almost always "No," if we expect the monitoring mechanism itself to be that further research. The establishment of new linkages between process and outcome can only be achieved, with any certainty, through carefully controlled and meticulously conducted clinical trials. It is unreasonable and dangerous to expect that every PSRO, or its analogue, can function as a research agency that tells us what good medicine is. One might with equal reason assert, as has been done in the past, that the best test of the usefulness of a drug is the sum of the judgments of individual physicians as they observe its effects on the management of their individual patients.

The uses of outcomes

Nothing I have said so far should be taken to mean that outcomes are not important in quality assessment and monitoring. Quite the reverse is true: Let me count the ways.

1. Outcomes, usually undesirable ones, can be used as a method of sampling or screening in order to increase the yield of process assessment by concentrating on cases with such outcomes. Perhaps the clearest example of this strategy is to be found in the "problem-status outcome" method described by Mushlin et al.[9] It is also visible as an important element in the Performance Evaluation Procedure (PEP) advocated by the Joint Commission on Accreditation of Hospitals.[10]
2. Outcomes can be used as a proxy for elements of process which are difficult to measure, or about which information is hard to get or is absent, provided the causal linkage between outcome and process is reasonably well established.
3. Outcome items can also serve as a supplement to monitoring process in order to ensure that important process elements have not been overlooked. In this case, they provide an added tier, so to speak, in the surveillance system.
4. Outcomes can also serve as a feedback mechanism that may lead to questions not only about whether certain process elements are

adequately represented in monitoring but, more fundamentally, about assumed relationships between process and outcome. As previously discussed, this latter is the point where monitoring and research are most likely to intersect.

5. The inclusion of outcome assessment can also serve to reinforce problem-oriented management, or "management by objectives."[11]
6. Attention to the attainment of prestated outcomes can motivate a more serious examination of process and be a powerful spur to reform. In Williamson's work on "health accounting," for example, one finds not only an implicit preference for management by objectives, but, also, a deliberate reliance on confrontation with failure in order to jolt physicians into action.[12]
7. Outcomes reflect the impact of all the elements that go into care, and of much else besides. They have, therefore, an integrative property which allows them to represent the contribution of all the health professionals to patient care. They also include the contribution of the patient to his own care. Unfortunately, this useful ability to pull together all these influences is also a weakness, if one wishes to explain how the observed outcomes come about.
8. Outcomes can not only include the patient's contribution to care, but also provide a means to enlist the client in the process of quality assessment and monitoring. The specification of the technical elements of process is an esoteric professional enterprise about which the client can only have rudimentary and, possibly, distorted knowledge. By contrast, the patient has a great deal to say about the interpersonal component in the process of care. He has at least as much to say about the outcomes he expects, to what extent these are attained, and how different outcomes are valued relative to each other and to the costs incurred in attaining them.

If I have been successful in my exploration of the role of process and outcome in quality assessment, one should never again hear a preference for one over the other, except by the poorly informed. But, I am also realistic enough to know how forlorn a hope this is. No doubt the debate will go on. In the meantime, it may be useful to present some general principles that govern the use of outcomes in quality assessment.

1. Obviously, the outcomes selected should be relevant to the goals of health care in any particular situation.
2. This also means that the outcomes must be achievable by good care and that the instrumentalities needed are available to, and under the control of, those whose performance is being assessed.
3. The outcomes must be specifiable in magnitude, frequency and timing. The paucity of information about the course of untreated and treated disease may make it impossible to specify accurately future outcomes, limiting their use in certain assessment schemes.[13]
4. The duration of outcomes as well as their magnitude must be taken into account. Hence the emphasis on a long-term perspective, often as long as a lifetime, in the measurement of health status.[14]
5. As a corollary to the above, one needs to consider the possibility of trade-offs between magnitude and duration for any one outcome, and among several outcomes. For example, a shorter life at a higher level of function may have to be weighed against a longer life burdened with greater disability.
6. As another corollary, accurate information on the relevant outcomes must be available and the outcomes must be subject to reasonably precise measurement. Since definitive outcomes may not appear except after long periods of time, information may be hard to get.
7. It is necessary to examine the consequences not only of taking a specified action but, also, of not taking such action, in order to get a complete picture. Thus, in evaluating the effectiveness of a surgical procedure, it is necessary to follow not only those who have had operations but also those who might have been candidates but were not.
8. The attainment of outcomes cannot stand alone as a measure of success. The means used in achieving these outcomes have also to be considered, unless it is assumed that resources are unlimited, which is far from true. This simple truth has only occasionally been appreciated in assessments of quality, as distinct from utilization review;[15] but it has received much attention in program evaluation, under the heading of cost-effectiveness and cost-benefit analysis.[7] In this context, it is useful to remember that the costs are not only in resources expended or risks taken; any assault on established values and social norms is also a cost which deserves serious attention, unless it is regarded as desirable and considered to be a benefit.
9. At an equally fundamental level, the problem of attribution has to be handled. Outcomes, whether good or bad, must be attributable, first, to medical care and, then, to the performance of those under assessment. Naturally, the longer the time that has elapsed between a specified activity and its consequence, the more opportunity there has been for the intervention of other factors, and the more difficult it becomes to assign responsibility for observed outcomes.
10. To conclude this decalogue, when adverse outcomes are used to assess quality, one must al-

most always examine the antecedent process of care to find out what went wrong, and how it might be corrected. The search for causes and remedies will often lead, even beyond process, to an examination of the structural characteristics that have encouraged or discouraged specified behaviors. The quality of care cannot be fully comprehended or successfully assured without understanding how structure influences process, and process influences outcome. No matter where one starts in this chain, one must ultimately deal with it as a whole.

8 I began this paper with a general frame of reference hoping that it would give meaning and perspective to the listing of research proposals; but, also, that it would generate a corresponding classification. Unfortunately, no truly satisfactory classification was found, because there is a great deal of overlap among the several categories that emerged. Nevertheless, I hope that the following is a reasonably orderly presentation in which the reader can discern the major features described in the introduction. There is also some attempt to have a very rough progression from studies dealing with basic concepts and measurement tools, to those that deal with implementation; but too much should not be made of this. Finally, there is the problem of specification. Obviously, it is not possible to be very specific in a general review like this. The text only sketches areas for research. But, whenever I could, I have cited examples of reported studies that might serve as more detailed models. In these the reader will find not only concrete embodiments of the more general descriptions in the text, but also, I hope, the raw material and inspiration that generates new research. Since the studies cited are meant only to be illustrative, the list of references does not serve as a systematic bibliography.

Basic explorations and studies of what constitutes quality

Specification and assessment by modeling As a first step in exploring what constitutes quality, at least in the technical sense, diagnostic and therapeutic management should be described and assessed as planned and sequenced activities or strategies. The models of such strategies that emerge will have a certain newness, although they use only existing knowledge and opinion. This information is obtained from reviews of the literature, from retrospective review of case records, from the opinions and values of expert clinicians, and from reasonable extrapolations from all of the above. The models that are constructed, in the form of algorithms or decision trees, are not only more precise and realistic representations of what is considered to be good care, but also provide the opportunity to test alternative strategies and to confirm or modify normative standards. Such formal testing is necessary because in these complex situations, the best course of action that intuition dictates may not be the best indicated by decision analysis.

The usefulness of this effort, then, is that it provides the basis for formulating criteria for assessing the quality of care. The models are also an important tool in medical education, as a vehicle for specifying and communicating to students the intellectual operations that constitute clinical judgment. Moreover, the attempt to construct such models reveals deficiencies in current information, identifies the most critical defects, and suggests the research needed to obtain the required information.

Illustrative examples of this approach to the exploration of diagnostic strategies may be found in the work of McNeil et al. on screening for hypertension[16] and on diagnosing pulmonary embolism;[17] and in the work of Neutra on the decision to operate for appendicitis.[18] Examples of the specification and testing of more complete strategies that include treatment as well as diagnosis are provided by Schwartz et al. for essential hypertension and renal artery disease;[19] by McNeil and Adelstein for hypertensive renovascular disease;[20] and by Tompkins et al. for acute pharyngitis.[2]

A review of these publications illustrates the importance of costs as a factor in the analysis. Nonmonetary cost in the form of risk is always a factor. In many instances monetary cost also figures prominently, for example in the papers by McNeil et al.[16,17] and Tompkins et al.[2] The problem of placing values on the consequences of various courses of action bedevils the analysis by Schwartz et al.[19] Methods for valuing the quality of life, in addition to its mere duration, are briefly discussed by Abt.[21]

Even a nodding acquaintance with this work suggests opportunities for further research, be-

yond its extension to cover a wider range of clinical situations. One line of inquiry is to go beyond determining how to best arrive at a certain diagnosis by studying the manner in which cases that present clinically identifiable problems can be assigned several diagnoses. The work of Ginsberg may serve as an illustration.[23] Another interesting study would be comparing the stepwise, logically parsimonious approach to diagnosis by the "shotgun" approach, which uses a very much larger number of tests, at least as a first step, to quickly narrow down the realistic alternatives. The possible gains in time and in the discovery of illness should be weighed against the monetary and nonmonetary costs.

Empirical testing of strategies of care Modeling, no matter how ingenious or creative, is limited by what is currently known or believed about the relationship between process and outcome. Additional knowledge must come from empirical studies of the elements of care or the more complex strategies of management. This may be accomplished by observing what might be called "natural experiments"; but definitive conclusions can only come from controlled trials, which can be regarded as an extension of research in clinical medicine. But clinical research is primarily concerned with the comparative effectiveness of alternatives; relative costliness, or the balance of costs and benefits, is seldom an issue. When clinical trials expand their scope to include not only risks and benefits to health, but also monetary costs and benefits, they address questions of social policy that are also essential ingredients in quality assessment.

The algorithms that currently define acceptable practice derive their authority from expert opinion. The work of Meyers et al. is one rather unusual example that is a reasonably direct empirical test of such an algorithm: one for the diagnosis of meningitis in children.[23] Some of the work cited earlier, for example that of McNeil et al. on the diagnosis of pulmonary embolism, can be regarded in the same light, since it uses patient records, though retroactively, to construct a diagnostic strategy.[17] A modest prospective study by Sturdevant and Stern that tests the ability of physicians to predict the finding of a stone on cholecystography is relevant because analogous estimates are used in decision analysis, when no better data are available.[24] An example of a much more ambitious clinical trial that tests consequences additional to medical outcomes may be found in the work of Piachaud and Weddell on the economics of treating varicose veins.[25] In this study the relative merits of injection-compression sclerotherapy and surgery were compared in a controlled experiment, using as criteria the condition of the limb at the end of three years, mortality from the procedure, the occurrence of immediate complications, the loss of patients due to nonattendance, the current monetary costs of the procedures, and their indirect costs as represented by travel time for treatments, days taken off work and loss of earnings. There was no attempt to construct a single measure for costs and another for benefits, but the advantage was clearly with sclerotherapy. The implications of such studies to quality assessment and to social policy are obvious and important, assuming that their findings can be accepted as valid.

Increments of cost and of quality As I tried to show in the introductory section of this paper, the relationship between increments of quality and of cost is a matter that deserves special attention. One way of studying it is through modeling and simulation. Using this method, McNeil et al. have estimated that it costs $8,485 to identify the last one out of 20 cases of pulmonary embolism in a young adult population with pleuritic pain, whereas the preceding 19 cases could be identified at an average cost of $595 per case.[17] Even more striking are the cost estimates of Neuhauser and Lewicki for a procedure for finding cases of cancer of the colon in a hypothetical population of 10,000 persons 40 years old or older.[26] The procedure is to first do a stool guaiac test and afterwards to x-ray those who show at least one positive test. What is varied is the number of successive guaiac tests that are included in the first step. Obviously, as the number of such tests is increased, the number of cases who show at least one positive test increases; more cases with cancer are detected, but there are also many more false alarms to be allayed by subsequent barium enema. Accordingly, the cost of finding an additional case of cancer rises steeply as the "quality" of care, judged by the number of stool guaiacs prescribed, increases; so much so that the cost of finding the additional case is estimated at an astounding $47 million when 6 tests are done, whereas it is only about $1,000 when one test is used as the screen, and about $49,000 when three tests are used. The values that are generated by these models are, of course, heavily influenced by the models' assumptions; and they could be inaccurate. My intent is to emphasize not the findings but the nature of the method and its utility.

As an extension of such explorations, under essentially hypothetical conditions, it might be useful to have clinicians formulate a set of increasingly stringent criteria of the quality of management, as well as an estimate of the gains in health expected from such stepwise increments of quality. Cost estimates could then be made, and the expected benefit compared to the cost. Hardwick et al. implemented what is, in effect, the first phase of such

a procedure when they asked house staff and general practitioners to specify the diagnositc tests they would perform for specified conditions (1) as an absolute minimum, (2) routinely, (3) if the case were to be presented at a medical grand rounds, and (4) if it were to be included in a research study.[27]

Needless to say, empirical studies are necessary to establish the relationships between quality and cost. A study of secular trends in the inputs into care for carefully specified conditions as compared to the outcomes of such care could be one initial approach. A good example is the study by Martin et al. of inputs into coronary care compared to the mortality during hospital stay.[28] While such studies are of great interest, contemporaneous comparisons of costs for similar services at different levels of quality may be more credible. In one such study, Jackson and Smith found no relationship between estimates of the quality of pharmaceutical services in community pharmacies and the charge per prescription.[29] Similarly, Schroeder et al. found no clear relationship between the cost of laboratory services ordered by residents and the assessment of the competence of those residents by their supervisors, even though there were wide variations in both estimates.[30] Rubenstein et al. have reported findings that suggest diminishing returns in outcome ratings with increments in the quality of the process of care.[31] These are only a handful of studies, but they show the way for more.

Studies of the client-practitioner relationship The research I have proposed so far defines quality in terms of the technical elements in clinical judgment. There is a corresponding universe of concern with the judicious management of the interpersonal relationship between clients and practitioners. Here, also, there are differing modes of interaction and varying styles and strategies of management that are more or less successful in achieving desirable outcomes. These outcomes include satisfaction, knowledge, change in attitudes and behavior, including compliance with recommendations. All these may, in time, be related to success or failure in altering health status. Fundamentally, there is little difference between studying the effects of two drugs on hypertension and a study of the effects of two ways of managing the interpersonal relationship on compliance with the drug regimen itself. Both are clinical trials, though they may draw on different bodies of theory and concepts. But, despite the similarity, the study of the effect of drugs is considered squarely within the domain of clinical practice, whereas no one is quite sure where the study of the interpersonal relationship belongs. Ideally, clinicians should be as interested in one as in the other, and actively engaged in research into both.

The literature pertaining to the client-practitioner relationship is too vast to permit a quick summary. A text book review by Bloom and Wilson serves as a useful introduction.[32] Lebow has reviewed the literature on patient satisfaction, discussed some of the problems of method, and made suggestions for further research.[33] A specific example of a clinical trial that examines the consequences of a change in the client-practitioner interaction is the study by Inui et al., which shows considerable improvement in the control of hypertension when physicians shift their attention from the manifestations of the disease to the behavior of the patients.[34] These encouraging findings can be placed against the discouraging experience in another clinical trial of the lack of success with either of two strategies, "augmented convenience" or "mastery learning," to improve medication compliance in hypertension.[35]

Views of what constitutes quality The research we have reviewed and proposed so far is expected to give a more precise answer to the question: What constitutes quality? Another approach is to study the opinions held on this subject by clients, administrators and professionals. A comparison should be made of the attributes that are believed to constitute quality and of their rankings in importance. The findings would be germane not only to quality assessment, but also to understanding problems in system performance.

An example that fits precisely in this category is the study by Smith and Metzner of the opinions of patients, physicians and nurses on what constitutes quality of care in prepaid group practice.[36] Marram's work on how much credence nurses are willing to give to the opinions of patients concerning different aspects of nursing performance suggests an interesting topic for further study.[37] The literature on patient satisfaction to which we have already referred includes much material from which inferences may be drawn about how clients view the quality of care. In this respect, patient satisfaction has an interesting dual nature: it can be regarded not only as an outcome of care, but also as a judgment by the client on the quality of care.

Continuity and coordination as attributes of quality There is a general presumption that the continuity and coordination of care are important considerations in the assessment of quality, even though it is not quite clear what these phenomena are, or how they fit into a formulation of what quality is. De Geyndt accords them a central place in the definition of quality by considering them to

be aspects of the "process" of care as distinct from its "content."[38] I have preferred to regard them as an aspect of the organization of what I have called "process."[39] More important than these formal distinctions is the precise definition and measurement of these attributes, and the study of their contribution to the outcomes of care. Empirical study should also include the degree to which existing differences in, or purposive manipulations of, certain structural features bring about differences in continuity and coordination of care. I would also like to see a test of the hypothesis that planned transfers of the responsibility of care from one physician to another might actually improve care by creating an opportunity to assess it anew, and by making the performance of physicians more visible among colleagues.

Important contributions to a more precise definition of the concept of continuity may be found in papers by Shortell[40] and by Bass and Windle.[41] However, the tendency to operationalize continuity to mean care by one physician or a single source of care, except by referral, does not seem to me to capture the essence of the concept of continuity. Almost all the studies I have seen have used this definition.[40,42-47] Bass and Windle tried to assess the relatedness of past to present care.[41] Shorr and Nutting defined continuity as the completion of a needed sequence of care.[48] Becker et al. describe a well-controlled trial of the effects of having children see the same physician at each clinic visit, and give a good review of the relevant literature.[49]

Description of prevalent patterns and strategies of care

In a previous section we discussed the modeling and testing of strategies of care. In this section we focus on a complement to the research described earlier, namely on the identification of how physicians behave in the real world, described in terms of the elements of care, as well as bundles and configurations of such elements. Here we seek to describe what goes on, what factors influence it, and what the consequences seem to be.

Studies of the elements of clinical behavior Much more information is needed about differences in the patterns of care among practitioners within a given setting and across settings. We need to understand what factors are responsible for such differences and what the costs and other consequences are. The factors involved could include personality attributes, knowledge, training and socialization, position within an organization, response to role models within the organization, and financial incentives.

Several studies of the use of laboratory and other diagnostic procedures can be cited as a good example of this type of research. The findings of Childs and Hunter suggest that monetary return on an investment in an x-ray machine may be a factor in recommending radiological procedures.[50] On the contrary, the persistence of large differences in the use of laboratory and other services in settings where there is no direct financial incentive to use such services,[51] or where the incentive is the same for all physicians, [52,53] suggests that other factors are equally or more important. Schroeder and his associates in a series of studies have looked into the correlates of differences in laboratory use and shown that these differences are not related to differences in competence,[30] or to differences in outcomes, or "productivity";[54] and that physicians may respond to a "cost audit" by reducing laboratory use.[55]

Identification of styles and strategies I have already defined a strategy as a plan for action, and discussed the importance of dealing with strategies in the definition and assessment of quality. A "style" may be defined as a habitual preference for certain modes of decision making which would manifest itself in components of strategies or strategies as a whole. For example, a physician may exhibit a persistent and pervasive preference for errors of commission over errors of omission, one manifestation of which may be a large redundancy in gathering information. He may give evidence of more than usually routinized or stereotyped behavior. These and other persistent yet undefined propensities require precise conceptual formulation and empirical study.

The notions of style and strategy have application beyond the solution of clinical problems. I have already suggested that they can also be used to study the client-practitioner relationship. They also apply to the way in which a practitioner manages an entire case load, hoping to achieve the most efficient allocation of his time, attention, and other resources among competing calls on them.

Styles and strategies can be inferred from physician behavior in real-life situations or under more artificial test conditions. Information on the rationale employed by the physician can be obtained more directly by having him explain, as he works, the reasons for doing what he does. Fattu has summarized some of the early work using this method, known as *réflexion parlée*, as well as other methods in studying problem solving.[56] A more recent example of the use of *réflexion parlée* in exploring clinical decision making has been reported by Kleinmuntz.[57] A study of the diagnostic process by Leaper et al. notes variations in the degree to which the clinical interview is either "stereotyped" or "adapted" to the problems of each patient, as inferred by an observer.[58]

Obviously, strategies that are identified through empirical study become candidates for testing, as described in the opening section on research proposals. I also believe that strategies used by "good" physicians are a more valid basis for the formulation of explicit process criteria than is the practice of having physicians list all the things that should be done for cases with a specified diagnosis.

Comparisons of norms with practice The literature teems with observations that physicians fall short of the normative standards of care. In some cases it has been suggested that physicians do not follow the standards which they themselves, as a group, have formulated.[59] Some have claimed that many errors in care are due to inattention by the overworked physician.[60] Others have shown that part of the deficiency in performance is due to lack of knowledge, while another part is due to not acting on what is known.[61] In my opinion, the explicit criteria lists which are often used to judge performance are, themselves, often faulty. One important inadequacy is that they fail to take account of the many contingencies, including multiple diagnoses, that modify the strategies of management. Moreover, it may be inappropriate to apply to office practice, especially to that of the generalist, criteria derived from strategies of management that are suitable in academic settings where highly specialized physicians are involved in the solution of difficult diagnostic and therapeutic problems.

Unfortunately, much of the above is only conjecture; and there is an urgent need for studies that attempt to understand why physicians conduct care in the way they do, before passing judgment that what they do is inappropriate.

Homogeneity and heterogeneity of performance An interesting question with many practical implications is whether physicians and other practitioners perform equally well across a range of activities and functions, or whether they do well in some and not so well in others. This involves examining the homogeneity of performance in the practice of individual physicians, as well as the ability of an institution to reduce variation in practice across physicians.

The relevance of this issue to quality assessment is most apparent when it comes to sampling. In one method, the "tracer methodology" proposed by Kessner, there is an explicit assumption that the performance of an entire system can be mapped by judicious selection of a small number of conditions that can stand for all the rest.[62] In many other studies, when a small number of diseases or conditions are selected for assessment, there is a similar assumption, even though it may not be explicit, or occupy such a central place in the design of the method.

Kessner has tested his assumption of homogeneity and found it, at best, frail.[63] Lyons and Payne have reviewed the literature and done further testing of the degree of compliance with normative standards of management by individual physicians, across diagnoses, in office practice,[64] and in hospital practice.[65] Intercorrelations tended to be low. There was, however, a suggestion that greater homogeneity might be found in the work of physician subgroups who have a more restricted domain. This clustering was also found in a study of the office care of a set of preventive and illness situations in children. Homogeneity of performance was reasonably high within each of these categories of conditions, and it was higher for pediatricians than for family physicians.[66]

The epidemiology of quality

Variations in the quality of care are not simply a random phenomenon. They are highly patterned and responsive to causative factors that we need to identify and understand if the quality of care is to be successfully safeguarded. The first step in this exploration is to answer the classic questions of epidemiological investigation: How much? Who? Where? and When? The results of these observations may suggest answers to the most critical of all questions: Why? The causal hypotheses that emerge could, then, be tested through more rigorous observational studies and confirmed by actual experimentation. But, for now, even the simplest of descriptive studies would add a great deal to the little we know.

On a larger scale, we can now say almost nothing about the quality of care for the nation as a whole, or for reasonably large populations in their natural habitats, other than what can be inferred from crude mortality, morbidity and utilization data. The one exception to this generalization that I know about is the study of a segment of care for the residents of Hawaii by Payne and his associates.[67] We are similarly in the dark about time trends: Is the quality of care improving, and how rapidly? This question is difficult to answer because it requires the separation of two phenomena: changes in the science and technology of medicine, and changes in the application of that science and technology. Both phenomena need to be assessed.

The epidemiology of quality can be viewed as manifesting itself in two populations: (1 the providers, and (2) the clients. Obviously, these are not two separate compartments. Variations in the quality of care received by clients are probably largely due to the kinds of providers who care for them. I

wonder to what extent the reverse could also be true.

The epidemiology of quality among providers The studies that might belong in this category overlap with those I have already described in the previous section on Description of Prevalent Patterns. If there is a distinction at all, in the earlier section I included studies that dealt with the detailed content of care and the rationale that explained differences in that content. Here, we are more interested in who provides care of good, bad or indifferent quality, and under what circumstances. Much of the literature of quality assessment deals with this question; but the restricted scope of most studies, and limitations in their design and analytic methods, makes generalization hazardous. It is clear, however, that performance is related to attributes of practitioners, attributes of the organizational settings in which they work, and interactions between the two. Among the attributes of the providers are education, training, specialization and length of practice. The role of personality characteristics including motivation, while suspected to be large, has not received much attention, except in studies of the academic performancle of medical students. Among the attributes of organizational settings, usually hospitals, that have been found to be influential, perhaps the most important has been affiliation with a medical school. Other attributes are involvement in residency and internship training, auspices, staffing, financing, size, and organizational controls on staff appointments and activities. Unfortunately, there is far from unanimity on whether these factors are influential and what their effects are. For example, we are still not certain whether the organization of physicians into prepaid groups results in better quality, and, if so, to what extent. The number of studies that have attempted actually to quantify the influence of each factor separately and in combination with others is particularly small. Whenever this has been done, what is most impressive is the very small amount of variance explained by the variables in the analysis, suggesting that we still know very little about the determinants of the quality of care.

The literature relevant to this section is so large that even a partial review would be a herculean task which I shall not attempt. Only some examples will be given. Among the earlier studies particularly outstanding is the work of Peterson et al.[68] and of Morehead et al.[69] Among the more recent work is the Hawaii study by Payne and his associates,[67] and the study of Medicaid beneficiaries in New Mexico by Brooke and Williams.[70] Of particular interest in the latter is the special attention to "outlier" physicians, those whose performance was markedly deviant.[70] Further studies of the characteristics of physicians that perform particularly well or badly could be a useful way of generating hypotheses about the determinants of performance.

Among the recent studies that have attempted to measure the magnitude of the effect of each of several variables on performance are those of Rhee[71] and of the Stanford University group responsible for the Institutional Differences Study.[72] Rhee used data from the study by Payne et al.[67] in which the dependent variable was a performance score based on compliance with explicit process criteria. Notable among his findings was the large effect of hospital characteristics compared to the effect of specialization and of organization into large, multispecialty groups; and, even more striking, is the magnitude of the unexplained variance. In their study on post-operative surgical mortality and morbidity the Stanford group not only found unexplained variance of a similar order of magnitude, but also failed to confirm the effect of factors usually thought to be conducive to quality, for example hospital size and university affiliation. Both of these studies can serve as models for future work. The Institutional Differences Study is particularly notable for the methods it developed to control for risk factors that influence the outcome of surgery and to specify and measure organizational variables. Its findings could perhaps be better understood if samples of records in the hospitals involved were to be subjected to an assessment of the process of care. I would accord such a study a high priority in this prospectus.

The correlations between organizational characteristics and performance are, of course, extremely important for system design. It would be interesting, therefore, to go one step further and determine whether the same physicians will alter their behavior when placed in different settings.

Epidemiology of quality among clients Unlike the emphasis on provider characteristics, there has been very little attention to client characteristics in studies of the quality of care. Although the receipt of quality care by disadvantaged populations is a matter of great social concern, and there is much public debate about it, the information bearing on the question is indirect. It deals mainly with differences in levels and patterns of utilization, in sources of care, and in morbidity and mortality data. Without minimizing the relevance of such information, it would be useful to have more direct and definitive assessments of the quality of care received by persons differentiated by age, sex, educations, color, income, occupation and other demographic and socioeconomic variables. Much of the differences would probably be related to differences in sources of care. But a question that

needs to receive special attention is whether the same institution, and the same practitioner, give different types of care to patients with similar medical conditions because of differences in the demographic and socioeconomic characteristics of the latter. Some adaptations of care to such characteristics are, of course, not only legitimate but, also, desirable and necessary. The issue is to determine whether the adaptations are made to maintain a high level of quality or whether quality suffers.

The literature having a bearing on quality of care for disadvantaged populations has been reviewed by Brook and Williams.[73] In another publication they describe their own findings in a study of Medicaid beneficiaries.[70] Lyons and Payne have described the relationship between age and the quality of care in their studies in Hawaii.[74,75] Kessner et al. have described the relationship between system performance and various demographic and socioeconomic factors, as well as source of care, revealed by an application of the "tracer" method in selected populations in the Washington, D.C. area.[63] Griner and Liptzin have described the effects of patient characteristics such as age, insurance status and ward or private accommodation on the use of laboratory tests in a teaching hospital.[76]

The identification of time trends is an important tool in epidemiological analysis that has been seldom employed in studies of quality. Hence, two studies that have information on this subject are particularly interesting. The first is a study of survival after cancer of the cervix uteri treated a decade apart. The findings suggest that improvements have been due mainly to a diffusion of knowledge from major centers to community hospitals, and to a much lesser extent to an improvement in medical science.[77] The second is a study of maternal mortality in Michigan from 1950 to 1971 showing that, in spite of spectacular declines in mortality, the percent of deaths considered "preventable" by the State's maternal mortality committee has increased markedly, from about 60 percent to about 80 percent.[78] A retroactive reassessment of the Committee's file of cases, applying current standards of preventability could be very revealing when compared to the contemporaneous assessments.

Most of the work on differences in the management of patients by the same provider, whether an institution or an individual practitioner, has been done in the field of psychiatric care. Perhaps the best known of these studies is the work of Hollingshead and Redlich in New Haven;[79] but there are many others. In a later work, Duff and Hollingshead showed that such differences in care can also be observed in a teaching hospital engaged in providing general medical care.[80] The understandable reluctance to look into this matter must be overcome. No program that provides care to clients of widely varying backgrounds can afford to ignore the possibility of discriminatory behavior in the application of care, in contravention of the most sacred traditions of the healing professions.

The relationship of structure to process or outcome

The reader will recognize that the epidemiological studies sketched out above often deal with observed relationships between structural characteristics (attributes of practitioners and institutions) and process, or outcome, or both. The more definitive verification of such relationships will require controlled experimentation. The major purpose of such studies is to safeguard and improve the quality of care. However, at the same time, they can elucidate the operational meaning of certain concepts, for example "continuity." To the extent that attributes of structure are found to be regularly related to performance, the more general use of such attributes as measures of quality will be more firmly established.

Development of basic tools for assessment

Many of the studies mentioned in previous sections, as well as many still to be described, cannot be done well unless certain basic tools are available. Thus, the refinement of existing instruments, and the development and testing of new ones, is a necessary part of research in quality assessment. At issue are the reliability, validity and cost of the basic instruments of assessment. A few of these, that I consider most important, will be selected for further attention.

Specification and measurement of outcomes. I have already indicated the ways in which the measurement of outcomes fits into quality assessment and monitoring. In this section, I shall describe briefly some ways of measuring outcomes.

One approach is to develop and use indicators f health and social well-being which permit a general oversight of a community of population. This approach is typified by the "sentinel events" proposed by Rutstin et al.[81] A useful area of research would be to specify appropriate indicators, develop methods for data collection, actually implement data collection and interpretation, and determine the usefulness of the information in bringing about change. Such a system would, of course, have to be adapted to the special needs of its users: whether a planning body, a program, or an organization that provides care. Any of these agencies may need to supplement information

which it collects with information from other sources, including census data and information from the National Health Survey. Naturally, the indicator conditions need not be only outcomes; a variety of process elements can also be included. Moreover, the system will be relevant not only to determination of quality, but also of need and unmet need, resource use, and so on. Many organizations are, no doubt, already involved in data gathering activities of this general kind. Perhaps the first step would be to review all that is being done, to document its current use, and to assess its potential usefulness. It is important to remember that a system overladen with data that are not useful, or are not used, can be as bad as one that has too little information. In any event, a pause for reassessment could be most helpful.

A second line to pursue is the development of "integrative" measures of health status that can represent the outcome of all the factors that influence health. The distinctive features of such a measure are that: (1) the impact of mortality and morbidity are combined; (2) morbidity is represented by a gradation of mutually exclusive categories of dysfunction; (3) dysfunction includes social and psychological, as well as physical, disability; and (4) the several dysfunctional states are weighted and summed into a single measure. The object is to arrive at a summary representation of the quantity and quality of life of a cohort of individuals over a period of time, often a complete life span.

Elsewhere, I have briefly reviewed the earlier stages in the development of this approach through the work of Sanders, Chiang, Sullivan, and Fanshell and Bush.[82] Since then, this area of endeavor has experienced an almost explosive growth which includes further work by Bush and his associates,[83,84] the work of Torrance,[85] and the work of Gilson and her associates on the Sickness Impact Profile.[86,87] Three collections of papers will provide the reader with a concentrated and reasonably current overview of the field.[88,89,90]

Perhaps the central problem in the construction of integrative measures of health status is that of valuation. There is need for empirical studies of the valuations placed by clients on different degrees and kinds of dysfunction. Expecially interesting would be differences in relative valuation by persons who are currently experiencing a particular level of disability. The effect of length of time in any level of dysfunction should also be examined, as suggested by Torrance.[85] Another line of development might be to try out a totally different method of valuation, comparison with a standard population, as proposed by Breslow and his associates.[91]

A third line of development is the construction of integrative measures of health status that are condition-specific. The global indices described above will probably be found to be lacking in sensitivity and specificity when used as measures of the quality of care. There is an opportunity to remedy these defects by developing analogous measures of the duration and quality of life of persons suffering from specific conditions, for example a particular cancer. The measures of function and dysfunction included can then be tailored to the condition, with attention given to including manifestations that can be prevented or remedied by proper care. The testing of such measures will also provide an opportunity to study the course of illness and identify additional outcomes that can serve as measures of quality.

A fourth line that might be pursued under the general heading of outcome measures is the development of condition-specific indicators of outcome. Here, only key indicators are used singly or in a profile; there is no attempt to integrate them, together with losses from mortality, into a single measure, as envisaged in the preceding section. But, obviously, there is a relationship between the two approaches, since the identification of individual indicators must be a step in the construction of an integrative measure.

The time elapsed since the institution of care is an important classifying variable in outcome studies. Accordingly, the indicators of outcome can be contemporaneous with care, or can follow care, in which case they are proximate (short-term) or remote (long-term). Short-term outcomes are of special usefulness in monitoring because they can be used to identify cases that require further study. This is a method of venerable ancestry, going back to the classic work of Codman,[92] and earlier. Its more recent manifestations in the work of Williamson and his students have already been noted.[12,9] A particularly useful model of the kind of research and development needed for constructing short-term indicators of outcome, is the work recently reported by Brook et al.[13] What still remains as an essentially unsolved challenge is the development of concurrent monitoring, using outcomes during the process of care. Of course, the conduct of care from day to day is constantly guided by such outcomes. The difficulty has been in adapting this everyday occurrence to a formal system of monitoring, other than direct supervision by peers or superiors. Some of the recent advances in computer-aided management do, however, suggest a possible solution.[60] What seems to be necessary is the constant feeding of selected information into a computerized system which raises an alarm when certain, prespecified, configurations of events occur or fail to occur during specified time intervals.

A final line of inquiry derives from adopting a definition of quality that includes phenomena such as satisfaction, attitudes, opinions, knowledge, illness behavior, and the like. This opens up the

whole area of methods in behavioral research which can be assessed and implemented by investigators having the necessary preparation. A model for such research that is closer to home may be found in the work of Hulka and her associates, first, in the development of a scale of client satisfaction and, then, in using that instrument to study its epidemiology.[93-95]

Improvements in the medical record The medical record is almost always the key document which contains the information for the assessment of care. Judgments of quality are heavily influenced by the nature of the record. There is also the possibility that the record is, itself, influenced by quality assessment activities.[96,97] Unfortunately, in spite of its key role in patient care and its evaluation, the record is often inadequate or poorly adapted to these purposes, especially in ambulatory care. The following are some proposals for remedying this situation.

16

While the record is often recognized to be incomplete, the accuracy of the information that it does contain is seldom questioned. The early work of Lembcke[98] is an exception to this generalization, as is the more recent work of Wiener and Nathanson.[99] It would be useful to test, by seeking independent verification, the accuracy of the history, physical findings, results of diagnostic tests, and so on. As a second step, it would be interesting to see what effect corrections of these errors would make on an independent judgment of quality based on a review of the record.

The completeness, and some aspects of the accuracy, of the record can be studied by arranging for independent direct observation of the client-practitioner encounter, or by recording it on videotape. Use of the latter method has been reported by Turner et al.[100], Zuckerman et al.,[101] and Steward and Buck.[102]

Alternative ways of designing records should be developed and tested as to their usefulness in the management of care as well as its assessment. Considerable work of this nature has been done in connection with the problem-oriented record. Examples are the studies of Tufo et al.[103] and Simborg et al.[104] Other work has been reported by Grover and Greenberg.[105] In work of this kind, the objectives include not only completeness and accuracy of information, but also ease in finding what is needed, and the ability to identify the practitioner's intent and reconstruct his rationale. Another objective might be the inclusion of the contributions to care of nurses, social workers and other professionals, so that assessments can be more inclusive. In a subsequent section I shall deal with the feasibility of even including entries by the patient himself.

At a more fundamental level, we have generally allowed the traditional content of the medical record to dictate what is included or not included in assessments of quality. In this way, the record constrains the definition of quality, allowing the tail to wag the dog. It seems to me that is is more reasonable to begin by defining quality independently of the record and, then, to design the record so that, alone or in combination with other specified sources, it can provide the information that corresponds to the initial definition of quality. In such an enterprise there would be a great temptation to demand an impossible degree of completeness. One must be adamant in resisting it. The more reasonable and challenging objective would be to define and implement the near-minimal set that would permit proper management and assessment.

I have already referred to the uses of computerized records in computer-aided management, and to the affinities between the latter and concurrent monitoring of care. Setting out to design such systems is an excellent opportunity to rethink the record and adapt it more suitably to its several uses. The relevant literature is immense. By way of examples, I have already referred to the work of McDonald[60] and of Schmidt et al.[96] Other work includes that of Wassertheil-Smoller et al.[106] and Barnett et al.[107]

The criteria and standards of quality No assessment is possible without some standard for comparison. In studies of quality, two or more providers may simply be compared. Another very prevalent approach is to judge performance by the extent to which it attains normative standards.

The urgent need to develop realistic and valid criteria and standards for condition-specific outcomes is implied in my previous discussion of the measurement of outcomes, and can be seen as a parallel activity. As Brook et al. have shown, the key steps are; (1) the identification of relevant outcomes; (2) ordering them by importance; (3) finding reliable and valid means to get information about the outcomes and to measure them; (4) specifying when during care or following it each outcome is to be measured, so that it is most discriminating (sensitive and specific) as a measure of performance; and (5) specifying the degree of progress toward each outcome that can be expected by good care, given certain attributes of the patient and his illness.[13] In addition to using such outcomes for retrospective review of care, there is need to develop methods for using them in concurrent monitoring, as has already been discussed.

Criteria and standards for the assessment of the process of care are very widely prevalent and are a basic tool in current assessment, monitoring and control activities. In spite of their central importance as a measurement device, the criteria lists have attracted little serious scientific analysis. I

shall devote the rest of this section to proposing ways of remedying this deficiency.

Perhaps the first step is to develop a taxonomy of criteria lists and similar formulations based on key attributes of their design and its underlying logic. Next would come an analysis of the possible implications of these features for quality assessment. The work of Rosenberg[108] is an example of an initial exploration in this direction. Much more work is needed, and soon.

As a result of the above, or independently, work should proceed on developing and testing alternative criteria designs. One way to go is to develop algorithmic formats that define more precisely optimal or acceptable strategies for management, taking account of frequently encountered contingencies. The work of Greenfield and his associates is an excellent example.[109] Initially, these algorithms derive their validity from expert opinion. Ultimately, they should be tested empirically, as indicated in an earlier section.

When a system of monitoring is designed, it might be useful to consider the use of several sets of criteria in stepwise fashion. For example, a simple list could be used for screening, with a more elaborate algorithm to be used for more definitive judgments in cases that fail the screen, possibly supplemented by a sample of cases that pass. A combination of a concise algorithm with judgments using "implicit" criteria may be tried out. The work of Mushlin provides an example of the latter.[9] Rubenstein et al. give an example of how criteria with "laundry list" and algorithmic components can be combined into a "decision index."[31]

As mentioned in the preceding section there is a mutual interdependence between recording and assessment. Hence, one part of the effort to develop and use alternative criteria designs is the work needed to redesign the record so that the criteria can be more efficiently applied.

The application of criteria lists to the assessment of process results, initially, in a "profile" of individual criteria that are met or not met. The derivation of an arithmetic average weights each item equally. Differential weights may be assigned to the several items and a weighted average obtained, as in the work of Payne et al.[67] These procedures lead to several difficulties. First, a given score can be obtained by different combinations of performance and non-performance of the several criteria on the list. Are these different combinations equivalent, or are there combinatorial interactions that are missed by simple summation? Second, we don't know what any given numerical score means. Is a score of 65 "good," "fair" or "poor"? Third, we do not know for certain what the basis for the weighting is, and how valid the weights are. Finally, a related matter, the construction of an average, weighted or unweighted, does not accord with the intuitive view that in some instances the absence of one critical element in care renders the entire care disastrous, no matter how many brownie points the care can earn in other respects. Of course, this could be handled by assigning near-infinite weight to such elements, provided the configurations that render them critical can be defined in advance.

The problem of weighting could be investigated through comparing the items on a criteria list with the corresponding algorithm, and to both implicit judgments of quality and the outcomes of care. Work that has a bearing on the question of weighting includes that of Richardson,[110] of Hoekelman and Peters,[111] of Lyons and Payne,[112] of Hopkins et al.,[113] and of Novick et al[114]

Another interesting line of inquiry would be to subject a set of records to assessment using different types of criteria. The cost implications of actually satisfying different types of criteria should also be determined and compared to the expected and, where feasible, the actual outcomes of care.

The social process, including group interaction, that leads to the formulation of, and agreement on, explicit normative criteria has been, to my knowledge, an almost totally neglected area of research. It would be useful to know how leadership is exercised, dissent handled and differences resolved. The effect of including health professionals other than physicians, and of administrators or, even, consumers should be studied. Similarly, the inclusion of physicians from a broader range of specialties, for example psychiatry, physiatry or public health and preventive medicine, might have an interesting effect. The content of the criteria as well as the process of arriving at them might be influenced. As a subset of such studies, it would be useful to look into ways of expediting the process of peer concensus and improving the decision by staff work that provides necessary information, forms or worksheets, and otherwise structures of the situation. Brook et al. demonstrate the usefulness of such staff work and, incidentally, comment on the impact of including a psychiatrist on the panel dealing with the outcomes of breast surgery.[13] We are indebted to these investigators for giving us this information. There are other workers in the field who also have considerable experience in such matters, but who have not taken the time to describe it for publication, perhaps because they have not realized how important it is. As a simple first step, I would suggest that this fund of information be tapped, even if it produces only descriptive accounts and informed guesses about what works and what does not

Monetary measures of costs and benefits The need to measure the monetary costs of inputs and the monetary equivalent of benefits arises frequently in assessing quality, as we have seen

already. Precise and valid cost and benefit measurements are also required to assess the effectiveness of utilization control and quality monitoring systems. In addition to devising rigorous and standardized cost accounting procedures, there are some important conceptual problems that require attention. The problem of arriving at monetary equivalents for nonmonetary costs and benefits has already been mentioned. In assessing the effectiveness of utilization control procedures it is possible to overestimate savings, and be unaware of shifts in the cost of care. Wyszewianski and I have indicated some of the ways in which this may happen.[115] A report by the Institute of Medicine provides a good summary.[116] Factors to be considered are that the days of care saved may be less costly than the average cost per day, that capital costs are not reduced proportionately to variable costs, that the hospital may function inefficiently if beds remain empty, that the physician may not be as productive in caring for some patients outside the hospital, that expenditures for care given outside the hospital in place of hospital care will rise, that these expenditures may not be covered by insurance, and that nonadmission to the hospital or premature discharge may generate future costs.

Specification and testing of system-design elements

Several activities already described fit under this new category of research and development projects. These include the design of alternative criteria formats and record systems. Selected additional features of system design will be described below.

Specifying the appropriate cut-off points in standards The determination of the appropriate cut-off points or levels in the standards for monitoring is a critical design element because both the yield and the cost of the monitoring effort are heavily influenced by that. There may also be other consequences, for example to the social acceptability of the system and dysfunctional adaptive reponses to it. An excellent example is the determination of the most appropriate "check points" for recertification of further hospital stay. To make this determination it is necessary to specify the factors that go into the analysis, to identify what information is necessary, and to implement the analysis, first, in model form and, later, in practice. Wyszewianski and I have indicated one possible way to proceed.[115] Averill and McMahon have offered a mathematically more rigorous model.[117] I consider the further development and testing of these proposals as matters of high priority.

Sampling and "enrichment" techniques There are two aspects of this subject that tend to overlap and become confused. The first has to do with the kind of probability sampling designed to obtain an unbiased picture of a specified universe of phenomena. Obviously, this is a critical issue in many assessment efforts; and much work is needed to develop efficient means of stratification and sampling. Some studies mentioned earlier, bearing on the heterogeneity and homogeneity of performance and the factors that influence it, would contribute to the knowledge needed to sample more efficiently.

There is another kind of selection which is not sampling in the statistical sense, but a method of screening. Its intent is to increase the yield for monitoring: to hit pay dirt, so to speak. Ideally, a method is wanted that would pick up every case that is managed suboptimally, while it excludes every case that is managed at an acceptable level of quality. In other words, a screen is wanted that is 100 percent sensitive and 100 percent specific. In the real world we, obviously, have to settle for something considerably short of this.

In considering further research in this area, let me begin by pointing out that not enough attention has been given to whether total coverage is necessary if a monitoring system is either to give a fair representation of performance or to be effective in achieving improvement in performance. Samples could be no less effective in achieving both these objectives. Pilot studies to verify this possibility would rank high in my list of priorities.

A fair amount of work has been done on developing selection or screening methods that are intended either to increase the yield from monitoring or, also, to bring about the most efficient separation of questionable from non-questionable cases. Examples are the work of Wolfe,[118] Riedel et al.,[119] Rubin,[120] and Glass et al.[121] Brauer briefly describes a variety of selection mechanisms and gives a longer account of an adaptation of the method described by Riedel et al. to reviewing psychiatric care.[122] Certain procedures that begin the process of review after observation of poor outcomes, such as the methods used by Williamson[12] and by the PEP system,[10] can also be regarded as concentration, enrichment or screening mechanisms. Despite all this, relatively little is known about the effectiveness of such screening techniques as measured by sensitivity and specificity ratios. Mushlin does provide a test of his scheme, which depends on reviewing the records of those whose original symptoms have not improved within a month of reporting for care. For

the three acute conditions tested, the sensitivity and specificity of this method are quite impressive.[9] By contrast, the method advocated by Riedel et al., which is essentially the selection of statistically deviant cases, does not seem to perform very well, judging by the preliminary reports that I have seen.[119]

One particular method of selection, that involved in the "tracer" method developed by Kessner et al.,[62,63] has been described in an earlier section.

Larger elements of design In this section, I want to call attention to the design of the assessment and monitoring endeavor as a total system in which there is a mutually supportive functional relationship among parts. I have tried to develop this idea in an earlier work.[123] Various expressions of it are daily seen in the design of a variety of systems including that of the PSRO program. The conceptual apparatus and the methods for testing such constructs will probably come from systems engineering and analysis, and will embrace both technical and social phenomena. Since systems analysis is another of the many subjects about which I know next to nothing, all I can do is to express the hope that it has something to contribute. If it does not, the work of designing and testing these larger systems will not stop, but the underlying principles that govern design and effectiveness will have to be formulated as the construction goes on.

Comparative studies of quality using different approaches

A great deal of insight into alternative methods of assessment can be obtained when they are applied to the same set of records, and the resulting estimates of quality are compared. An excellent model is the early work of Brook,[124] particularly when studied in its more detailed presentation.[125] More recently, Brook et al. have proposed that separate sets of process and outcome criteria be developed simultaneously for a number of conditions, with subsequent comparison of the ratings of quality accorded to the same care using both sets.[13] The primary purpose in such studies is not to find new facts about the link between process and outcome, but, given existing knowledge, to determine whether the proper formulations of process and outcome criteria have been made. However, when discrepancies between rating based on the two sets are found, they could lead to questions about the validity of what was thought to be known about the process-outcome relationship.

Further development of promising current approaches

In recent years, a number of promising new approaches have been developed and tested more or less rigorously. In most cases, a great deal remains to be done. To select one or more of these approaches and invest in further testing could be very rewarding, since it would build on existing work and benefit from the advice and collaboration (and, sometimes, the afterthoughts) of their originators. In fact, many of the proposals for research that I have included in this review are derived from a critical examination of what has been published about these new approaches.

At the risk of some repetition, let me mention some of the approaches that I think are most promising. The order in which they appear is not intended to signify either perceived importance or personal preference.

Outcomes as measures of quality. The major recent exemplar of this approach, the Institutional Differences Study, demonstrates both its usefulness and its limitations.[72] The fundamental assumption in this, and similar, studies is that when statistical corrections are made for known risk factors that influence outcome, a great deal, if not all, of the variation that remains is accounted for by differences in the quality of care. I believe that a direct test of this assumption should be attempted by independent assessment of the process of care, using a variety of methods. I suspect that the resulting correlation between process and outcome would be low, suggesting the need for better methods to standardize for risk factors, as well as a more fundamental approach to process assessment. In particular, the decision to operate or not was not subjected to assessment in the Institutional Differences Study, as the investigators take pains to point out. I suspect that many of its anomalous findings are traceable to this basic weakness.

The method of assessing proximate and intermediate outcomes of care, so well described by Brook et al.,[13] is another promising approach that requires and deserves much further development along the lines indicated by its proponents. This includes extension to other diagnostic categories or conditions; actual implementation of the criteria already developed to test feasibility, accuracy and cost; and comparison with simultaneous process assessments.

Outcomes as cues and motivators for process assessment There is no sharp line of demarcation between the use of outcomes as measures of quality in their own right and as screening devices to select cases for process review. In fact, my attribu-

tion of primarily one function to some methods and another function to other methods is likely to be challenged by their originators and advocates. However, I do perceive the approach developed by Williamson[12] and used by Mushlin,[9] as well as some basic elements in the PEP method of the Joint Commission on Accreditation of Hospitals,[10] to fall more comfortably in the category of monitors of outcome rather than assessments of quality based on outcomes. The method described by Brook et al.[13] appears to be intermediate between the purest forms of the two classificatory categories that I have proposed. But none of this discussion on classification need deter us from noting the potential usefulness of these methods and investing effort in their further development. I am particularly impressed by the simplicity of Mushlin's approach, and its great success in separating significantly deficient care from acceptable care. Whether this will remain true when additional conditions are tested remains to be seen. As to the PEP approach, it offers many opportunities for further development, for example the determination of responsibility for complications that occur in the hospital and the inclusion of some outcomes that appear after discharge from the hospital, as proposed many years ago by Codman.[92] To return to the distinction which forms the basis of my classification, if these methods to indeed use outcomes as monitors and screening devices, rather than as more complete representations of quality, a clear recognition of this distinction could lead to simplification of the measures of outcome and to their assessment in terms of their screening efficiency. I believe that this would be a very useful and important development.

The occcurence of preventable adverse events Rutstein et al. have recently reminded us of the potential usefulness of this time-honored method of monitoring the health care of a population.[31] In this method, attention focuses on outcomes and other events that are preventable, at least to a significant degree, when good care is available and is used. Obviously, this is little different, in principle, from the approaches described in the preceding section, except that we are now speaking of populations rather than of patients. A historically significant method of quality assessment and control that probably belongs under this heading is represented by the activities of the maternal mortality and perinatal mortality committees originating in the landmark studies of the New York Academy of Medicine.[126,127] Since then, much information of this kind has been accumulated over long periods of time in many states and localities. The usefulness of this historical material is demonstrated by a recent analysis of maternal mortality data in Michigan.[78] A current application using infant deaths in hospital, has been reported recently from England.[128] The key feature of these studies is that the problem of attribution is handled by a case-by-case analysis that leads to a determination of whether there were preventable or avoidable adverse circumstances and by assigning responsibility for these. Broader application and testing of this method, using a wider range of conditions, seems to be well worthwhile.

The occurrence of preventable progression of illness or disability This is another member of the family of outcome-oriented methods which cannot be fully differentiated from some of its companions, except in emphasis. It has particular affinity to the preceding category, Preventable Adverse Events, except that, in this instance, the focus is on the preventable progression of illness from an earlier stage, which is presumably more amenable to treatment, to its later, more advanced and recalcitrant manifestations. This approach owes its more recent saliency to the work of Gonnella and his associates who refer to it as the "Staging Concept."[129,130] In one way, the staging of disease creates more homogeneous categories, so that the attainment of outcomes can be compared with greater confidence that differences are attributable to care. However, Gonnella et al. also argue that the stage at which a disease comes under care, either initially or at some later date, tells us something about earlier access to care and the quality of that care, if care is provided. As in all outcome studies, the problems of interpretation are many, but this approach does simplify population monitoring to some extent because the necessary data can be obtained from within the patient population. For example, a hospital may be unable to precisely identify the population it serves or to obtain useful population data, but it can characterize all, or a sample of, its admissions as to stage of illness and preventability of progression to such a stage. Further empirical testing is warranted.

Indicator conditions: "trajectories" and "tracers" There is a large number of studies in which one or more conditions are selected and the career of patients with these conditions is followed as the patients proceed through the system. This could be call the "trajectory" approach, since the emphasis is on what happens at each successive step in a progression that is, too often, a tragic Odyssey of accumulated failures. Examples are provided by Brook and Stevenson,[131] Starfield and Scheff,[132] Novick et al.[114] and Shorr and Nutting.[48] The "tracer" method developed by Kessner et al. can be seen as a highly systematized selection of such indicator conditions, each with its distinctive trajectory. The systematizing or organizing device

is a prior conceptual mapping of a field and the purposive selection of conditions to represent all the major elements in that field.[62,63] Another characteristic of the "tracer" method is its emphasis on combining population and patient data to achieve an epidemiologic investigation of the problems of medical care. However, this epidemiological perspective could be incorporated in the "trajectory" approach, in which case each trajectory becomes a tracer (more accurately, the path of a tracer). Such semantic games aside, there is much opportunity for further work in this area. I particularly like the notion of a planned selection of tracers, with a view to systematically sounding the corpus of medical care in its totality and to include both patients and non-patients. However, since this requires a massive effort, more modest and circumscribed applications should be tested first.

Second surgical opinion programs An idea that has captured a wide audience is the possibility of controlling unnecessary surgery through either making available or requiring second surgical opinions.[133] This is a particularly interesting approach since it is truly preventive. There are many opportunities for research and development here, including tests of the reliability and validity of multiple surgical opinions, acceptance by clients under voluntary and mandatory systems, acceptance by physicians, effect on relationships among physicians and surgeons in a community, effect on initiation of recommendations for surgery, effect on the client-physician relationship, and so on. To answer the question of validity, long term follow-up under controlled conditions would be necessary. Such studies are now in progress.[134]

Computer-aided management I have discussed earlier the affinities between computer-aided management and several facets of quality assessment, including the design of records, the identification of critical events, and concurrent monitoring. I see this as an area of much fruitful further development

Integrative measures of quality

I have already said that most studies of quality focus on a relatively small segment of care provided by one professional, and that there is need for more "integrative" measures that include the contributions of several professionals during complete episodes of care and sequences of such episodes. I have also pointed out that outcome measures, especially those that are more inclusive in extent and duration, are by their very nature, integrative. The approaches that use the "trajectory" or "tracer" methods also have an integrative property since they often include outcome as well as process measures, and follow the course of care through successive stages, levels and sites, so that they reflect the cumulation of deficiencies at each of these junctures.

There is need to develop methods that incorporate process and outcome elements that are especially selected to represent the contributions of the several professions that are involved in patient care, because these elements are particularly responsive to the contributions of each of these several professions.

Another way of taking a more complete or integrated look at performance is to use as a unit of analysis the entire case load of a practitioner or institutional provider, so that the assessment includes not only the adequacy of care, but also the optimal allocation of resources among cases.

Applications to special areas

On the one hand, there is need to develop integrative measures of the quality of care. On the other hand, there is need to adapt assessment methods to special populations, to categories of care, to particular professions, and so on. This is especially true when an agency or organization that provides patient care includes very large elements of very diverse programs, including general ambulatory and inpatient care, long term care, physical rehabilitation, care for mental illness and alcoholism, etc. Within each of these categories, care could also be differentiated according to profession, for example: pharmaceutical services, dental services, nursing care, social work, dietetics, physical therapy, radiology, pathology, anesthesiology, and so on. It is impossible for any one reviewer to encompass the literature in all these specialized areas. I shall not even try to offer a partial description. However, the reader seeking an introduction might find useful some selected items that have come to my attention.

Freeborn and Greenlick have attempted to systematize approaches to the assessment of ambulatory care.[135] Christoffel and Lowenthal have reviewed the newer methods that are available.[136] A publication of the American Society of Internal Medicine provides a recent anthology on ambulatory care assessment.[137] The American Nurses Association has performed a similar function for methods of assessing nursing quality.[138] A collection of papers in the quality of pharmaceutical services can serve as an introduction to that very interesting and important area of application.[139] More will be found in a publication by Knapp and Smith.[140] A collection of papers on the assessment of dental care, though not so recent, could also be

helpful.[141] Much information about the assessment of mental health services in general, together with detailed information about one approach, will be found in a book by Riedel et al.[142]

Consumer perspectives and the consumer's role

There are many who are distressed by the near total domination that the professionals have exercised over quality assessment, from its deepest roots to its most slender branches. And yet, it has been very difficult to involve the consumer in quality assessment in a meaningful way. Aware of this problem, The Institute of Medicine has identified consumer participation as one of five "priority areas for quality assurance," and has devoted considerable space to it in its "research agenda."[116] In this section I shall deal with some ways in which consumers can participate in quality assessment, hoping that I will stimulate further thought, research and development.

Representing client values in the definition of quality At the root of quality assessment is a view of what attributes of care constitute quality, which imprints itself on everything that follows. It is, therefore, important, as I have argued earlier, that the clients' values, preferences and expectations be included in the definition of quality from the very first.

The choice of outcomes as a measure of quality is, in itself, a means for making the notion of quality easier for the patient to understand, and closer to his concerns. But not all outcomes are equally comprehensible or relevant. Outcomes that are defined in terms of physical or social function are much more meaningful than clinical, physiological or chemical measurements. The valuations that are placed on alternative outcomes, when there is a choice, could differ not only between clients and professionals, as groups, but also among clients as individuals. All these speculations are subject to empirical study, as is the degree to which the conduct of care takes account of client values and preferences, collectively or as individuals.

As we said earlier, clients also have preferences with respect to the process of care. Generally, these focus on the management of the interpersonal relationship. One cannot emphasize too strongly that in this one aspect of care the patient is fully as expert as the professional, if not more so. In fact, a good case could be made for having the client set the criteria of good care in this regard. Clients also have views about the technical component of the process of care. Sometimes, these are of debatable validity, as when an injection is demanded or, even, a hysterectomy. At other times, the expectations of the well-informed client can be quite valid, for example when more complete prenatal care is demanded or too ready use of some x-ray examinations or of antibiotics is resisted. In any event, it would be interesting to study the influence of such expectations on professional performance.

I have already commented on the dual nature of client satisfaction: as an "outcome" of care, and as a judgment on care. In the latter instance, it could be seen as the patient's estimate of the quality of care. It is remarkable how infrequently the patients' estimate of the quality of care he has received in a defined instance has been compared to that of a professional estimate of the quality of the same care. The only example that I can think of was reported by Ehrlich et al. many years ago.[143] This is an obvious area for further study.

Clients as sources of information For some data used in assessment the patient is the only authoritative source: for example, data on the patient's knowledge, opinions, satisfaction and the like. For other data, the client is an alternative or verifying source: for example, concerning services received and assessments of function or disability. There is much room for research on ways of obtaining reliable and valid information at low cost, and its incorporation into quality assessment and monitoring activities.

It is perhaps ironical that "the patient's record," in any medical setting, is not only totally barred to the patient, but also contains no direct entries by the patient concerning how he feels, what he knows, or how he perceives his care. It would be a fascinating experiment to see whether it would be possible to have the patient make entries into the record, either in narrative form or as checkmarks on a list of questions. It has been suggested that patients might be unwilling to express negative feelings for fear of reprisals. This is subject to testing and, if verified, entries might be made on a separate document that only later becomes part of the record. Even the answers to the question, "What was the one best thing that happened to you today?" would be most revealing.

Participation in monitoring Participation in monitoring can perhaps occur indirectly and informally through the manner in which the well informed patient responds to physician initiatives and participates in the client-practitioner interaction. This is a matter that can be studied.

A specific example of this more general category is the degree to which patients participate in second surgical opinion programs and how they react to non-confirmation of the initial recommendation. There is already evidence that participation

in voluntary programs is not high and that patients do not always act in accord with a second recommendation, whether it is confirmatory of the first or not.[133,134] Attempts could be made to alter client behavior and to test their success.

Some have argued that the paient's medical record should be accessible to him at all times. Shenkin a d Warner have proposed that this should be required by law, and have speculated on the possible advantages and disadvantages that might ensue.[144] Stevens et al. described experience with an actual trial which suggests that at least some patients can participate usefully in monitoring their own care, while others either cannot or are unwilling to try.[145] More work is called for.

It is not clear to what extent patients can participate in a more structured and formal manner in monitori the care they themselves receive, but there are many opportunities to find out. For example, Bouchard et al.[146] and Burger[147] have reported on experience with involving the patient in auditing the problem-oriented record by having the physician's evaluation sent to the patient for comment.

Membership in audit and utilization review committees As far as I know, the inclusion of clients on audit and utilization review committees is virtually unexplored territory. Goldblatt et al. describe professional resistance to inclusion of consumers on such committees in a project on mental health evaluation.[148] But, they also note very briefly that a "consumer opinion subcommittee" that "independently investigated clients' opinions about care and analyzed complaints of referring institutions and practitioners" influences selection of topics as well as cases for review. In this way, "consumer dissatisfaction could be turned into criteria for good care."[148] In the same way, consumer participation on grievance committees and similar bodies could be linked to the activities of quality assessment. There is urgent need for careful trials of a variety of mechanisms for involving consumers directly and indirectly in quality assessment and assurance.

Patient contributions to care The importance of the client-practitioner relationship to quality assessment has been a recurring theme in this paper. The care of the patient is a joint enterprise which includes contributions by the patient as well as the professional. Quality assessments of care may give the professional too much credit for success or too much blame for failure unless careful attention is given to the role of the patient. But, there is much more involved than simply deciding who is responsible for what. The assessment of the patient's contribution to his own care, and study of the factors that influence that contribution, is a legitimate and important area of research in its own right.

Quality assessment and monitoring as a social process

Perhaps the most difficult problems in establishing quality and utilization control mechanisms, and in operating them effectively, are social rather than technical. It is of the utmost importance to redress the imbalance in past and current research by paying at least equal attention to quality monitoring and control as a social process. In my earlier work, I have described briefly the kinds of factors that provide a context of quality assessment and monitoring.[149] I have also speculated on the nature of the factors that influence the effectiveness of monitoring and review.[123] More recently, Freidson has presented an incisive analysis of the social process in the implementation of PSROs "as part of a larger class of issues connected with the social psychology of work and its control."[150] Additional speculations, from a more operational viewpoint, can be found in a series of comments on a paper by Morehead,[151] and in a paper by Bellin.[152] Jacobs et al. describe and discuss factors in the implementation of the PEP system of assessment developed under the auspices of the Joint Commission on Accreditation of Hospitals.[10] Goldblatt et al. give an excellent account of experience with utilization review committees in a collaborative project in mental health evaluation.[148] Using these sources, and others like them, it would not be difficult to formulate a series of hypotheses as a starting point for research. But there remains a great and urgent need for an approach to the study of the implementation of quality monitoring that rests on a sound and systematic conceptual base. The development of this conceptual foundation is, itself, an area of scholarly research. However, it is likely that no unifying framework will emerge, and that this social phenomenon, like all others, can be seen and understood using a variety of theoretical constructs that are provided by economics, political science, sociology, organizational theory, anthropology, history, law, and so on. My pessimism notwithstanding, the student of medical care organization will continue to hope that, in some way, relevant contributions from the theoretical treasure of all these disciplines can be brought together coherently, so he can understand fully the problems that he faces.

At the broadest level, there is need to understand what might be called the politics of quality monitoring and control. This could include the role of government, of the organized professions, of the organized constituencies, including the

health insurance sector. The interrelationships between regulatory government agencies and the organized profession is a particularly critical area for study. All these forces need to be understood as they interact at national, state and local levels.

At the level of the hospital, and similar institutions, it is important to understand the power relationships within the hospital as well as those between the hospital and its environment. The organization of physicians in the immediate environment of the hospital and within it is a particularly critical element. Comparative studies of hospitals that have different goals and/or are differently organized could be quite revealing. Such studies might include comparisons of government-owned to community hospitals, and of hospitals that emphasize teaching and research to those that confine themselves to patient care.

The local PSRO functions as a key link between the hospital and the organized profession in its environment, and between the latter and governmental regulatory and financing agencies, as well as other third party payers. The dynamics of PSRO operations should become, therefore, an object of intense scholarly scrutiny. Unfortunately, it is quite likely that this will be countered by an equally intense determination to avoid being studied. Nevertheless, careful study of the PSRO as a social organization is absolutely necessary, and openess to such study should be a condition for formal recognition.

In this progression from larger to smaller social units the next critical level is that of the audit and utilization review committees, where much of work of assessment and monitoring is done. The structure and roles of these committees should be studied in differently organized hospitals with differing linkages to the local PSRO. The dynamics and effectiveness of the committees are likely to be influenced by the structure of the committees themselves, including variations in who is represented on the membership roster, and how the committee is linked, structurally and functionally, to centers of power and influence in the hospital as a whole. In an earlier section, I mentioned the importance of studying the dynamics of developing the criteria and standards that operationally define quality.

At the most disaggregate level of analysis one finds, as always, individuals who hold key roles, and who carry on their shoulders, as it were, the burden of the most massive of social enterprises. Depending on whether these persons are seen as prime movers or mere pawns, their behavior either determines or reflects the manner in which the total enterprise ultimately functions. In this instance, I see as if at the focal point of some giant lens turned to the sun, the lonely figures of the "nurse coordinator" and medical adviser." If I had a choice, I would go first to these, and try to understand how they work with each other, and how each related structurally and functionally to their respective peer groups, to the internal utilization and audit committees, the clinical department chiefs, the chief of staff, and the hospital administrator. There are usually others who are important as leaders, though informally; and still others who, like the pathologist, can have disproportionate influence because they control information critical to the assessment and its conclusions.

And, finally, what of the individual physicians themselves, who are the object of this seemingly unrelenting scrutiny? The manner in which they respond individually, through their formal and informal groupings within the hospital, and through their many-tiered and inter-locking professional ogranizations, including the PSRO, determines whether the program achieves its objectives, or whether it is so watered down and subverted that it becomes a ponderous and costly apparatus, making a brave show, but achieving little.

Effectiveness and the factors that influence it

Finally there is no way of escaping the most momentous of all questions: whether quality and utilization review activities are effective, and sufficiently so, to justify their immense social and monetary cost. It is remarkable that even now, when we have made a political decision to construct the awesome machinery of the PSRO, the answer is that we do not know. Some years ago I reviewed what was then known about the effectiveness and costs of quality and utilization review mechanisms.[123] Very recently, a committee of the Institute of Medicine reassessed the situation not only by reviewing the literature but also by obtaining information directly from operating programs.[116] My conclusion was that in some instances there was success, whereas in others there was failure, and that we did know for certain what accounted for either, though one could speculate at length on the matter. I understand the conclusions of the Institute of Medicine to be not much different. It is still not clear what hospital medical audits accomplish, if anything. By contrast, the utilization control programs of hospitals do occasionally report savings, but these tend to be overestimated because of improper accounting assumptions. Ambulatory care claims review, where studied, has been cost effective, but this is mainly or entirely due to the administrative component, as distinct from professional peer review. All these "savings," when they do occur, are to the fiscal intermediaries. The social costs can be shifted. As to the effect on the health of people, almost nothing can be said.

There are many reasons for this rather dismal state of affairs. The most fundamental are the

absence of sound conceptualization of the problem to be addressed, the deficiencies in the basic tools of measurement, and the absence of soundly designed and properly controlled studies. No accumulation of case reports from operating agencies, no matter how lengthy, can resolve these problems. Case studies do, however, have the virtue of generating hypotheses for further testing.

In subsequent sections. I will suggest some areas for study, without trying to cover the entire range of possible research. It will become obvious to the reader that there is much overlap between these proposals and others that were made earlier under different headings. In particular, there is a close tie between the specific considerations that come up in this section and the issues that were very broadly sketched in the preceding section on Quality Assessment and Monitoring as a Social Process.

Changes in physician and client behavior It is important to document the changes that occur in the behavior of physicians and other practitioners as a result of instituting quality and utilization review mechanisms. Rather simple before-after studies are useful; but, where possible, contemporaneous controls should be provided. In these studies, it is necessary to keep in mind, and look for certain "dysfunctional" behaviors, such as a tendency to lengthen stay up to the "checkpoint," or to discharge prematurely, as a response to a certification program; the likelihood that physicians will "manage by criteria," resulting in many redundant procedures; and the possibility that evasive actions will be take, for example by using a different diagnostic nomenclature or by moving patients to other settings. It is safe to say that everything that human ingenuity can devise will be used to tame a regulatory mechanism, and the researcher must be prepared to anticipate and study such behavior. Some attention to possible adverse or unintended effects can be seen in an interesting paper by Brian on the impact of a utilization control program in California.[153] Brian concludes that the program was effective without evidence that needed care was denied or costs shifted to others. In direct contrast, two reasonably well controlled studies of a hospital-stay recertification program in Pennsylvania showed no effect on hospital use even though the state Medicaid agency was much impressed by the reduced rate of unjustified stays.[154,155] A reasonable, though unverified, explanation of this discrepancy is that the program did not alter hospital use, but did improve the documentation needed to obtain payment for care.[154] If so, it is only proper to ask whether this new documentation is a better representation of the truth, or only a more credible distortion. To the extent that it is the latter, we may, as a society, be turning out the most expensive fiction every written!

I included the client in the title of this section because of a conviction that any change in the behavior of the practitioner is likely to have its counterpart in the client. Almost no consideration has been given to client behavior in studying the effects of quality and utilization control mechanisms, and I am not sure that there will be any, unless considerable costs are shifted to the client. Nevertheless it is interesting to speculate whether patients may be more likely to patronize physicians who are less scrupulous in accepting the strictures on hospital admissions and length of stay. Might the patient develop symptoms when informed that the approved hospital stay is about to expire? Could the local hospital lose community support for seeming to repeatedly refuse admission or to throw patients out?

Effect of technical design characteristics I have discussed in an earlier section some technical design elements that might influence the performance of the quality monitoring system. The relationships to effectiveness could be inferred from observations of existing variants or tested by intentional manipulation under controlled conditions. Examples include studies of the effects of different criteria formats, of varying the hospital stay recertification checkpoints, and of testing the cost-to-yield ratio of alternative sampling and enrichment schemes.

Intra-institutional "social design" characteristics This category subsumes a very large area much of which I cannot see clearly. I will mention only a few of the more obvious kinds of studies.

It seems to me that the legitimacy of the quality monitoring effort in the hospital, and the degree of commitment to it by its key figures would be a very important factor in the effectiveness of this effort. The written and verbal declarations of board members, administrators and physician leaders would be one source of direct information. Perhaps more valid would be the inferences drawn from how the assessment and monitoring effort is structured and how it functions. Much can be learned from an examination of who chairs the audit and utilization review committees, and who the members are. It is also important to know what decision-making power the committees have, how their recommendations reach the executive echelon in the hospital, and the degree of influence the committees have on the key centers of power in the hospital.

The legitimacy of the criteria and standards incorporated in the monitoring system could be a particularly important factor. One significant variable is whether the criteria are externally imposed, developed internally or a mix of the two. If the

criteria are, at least to some extent, internally developed, the degree of participation in criteria formulation may influence the adherence to the criteria by the physicians as a whole, or may differentiate between participants and nonparticipants with regard to adherence. If the criteria are externally sponsored, the identity of the sponsor could be important: for example, whether it is an insurance carrier, a government agency, the local PSRO, the Joint Commission of Accreditation of Hospitals, the American Medical Association or one of the specialty societies. In a large system such as the VA, sponsorship by the central office versus the local institution could be a differentiating variable.

The structure of sanctions and incentives as it operates on individual practitioners cannot fail to have an important effect. One feature is whether the monitoring system confines itself to studies of patterns of care or whether it goes beyond that to identify individuals whose practice is called into question. The methods used to communicate and interpret findings are determined to a large extent by the initial decision to identify or not identify individuals; but, irrespective of that, there are many options. Newsletters, general staff meetings, and meetings in smaller departmental or subdepartmental units could have different impacts. It is also important who is responsible for communicating the information; the nature of his involvement; and whether that person meets with the entire staff, a small group, or a single individual.

More important still is the manner in which findings about performance become instrumental in influencing the careers of individual practitioners. It may be that the major consequence of nonadherence to standards is approval or disapproval of payment by an insurance carrier or government program. While important, occasional brushes with a third party payer may not be as effective in influencing behavior as would be the certain prospect that the physician's performance record will be considered in determining his practice privileges, promotions, access to prestigious appointments, and other organizational rewards. The use of rewards and their possible impact should receive serious attention. Would it be possible, for example, to grant priority in admission to the patients of physicians who have a consistent record of very few unnecessary admissions or stays? While the absence of a reward is, itself, a punishment, it could well be that there is a significant difference between a system that focuses on punishing wrongdoers and another that stresses rewards to those who have an excellent record of performance. Finally, any incentive system will be inoperative if knowledge about it is not shared and if its certain and impartial implementation is in question.

Poor performance is not always primarily attributable to individual failure. Quite frequently there are organizational problems that interfere with good work by practitioners. In that case, and often when individuals are at fault, it is necessary to make changes in the organization. This requires the full support of those who hold executive power. Hence, the manner in which the quality monitoring system is linked to the executive, and the nature of its influence at that level, become critical elements in studying the factors that modify performance.

Education, either alone or in conjunction with sanctions, features prominently in attempts to bring about change in the behavior of practitioners. The variables that are likely to influence effectiveness include, first, the relative emphasis placed on education as compared to sanctions. Then, there are different educational strategies that could be more or less effective. One distinction that has been the subject of much speculation is that between an educational program based on topics of general interest and one that is guided by audit results. The latter approach can be highly individualized by tailoring continuing education to the deficiencies in an individual's performance. It is also claimed that participation in audit and review activities is itself educational and helps motivate change in behavior. All these speculations, as well as other hypotheses about the differential effectiveness of alternative educational strategies, are open to testing.

Strategies of client education are also a relevant variable, since patient cooperation may be an important factor in achieving the objectives of quality and utilization review.

Supra-institutional "social design" characteristics The sanctions and incentives inherent in a quality monitoring and control system may act on the institution as a whole, in addition to their effect on individual practitioners. The interests of the institution are likely to have great impact on how committed it is, as a collectivity, to quality and utilization control. For example, it is generally believed that when beds are plentiful, neither the institution nor individual physicians are motivated to keep patients out. Self-interest works in precisely the opposite direction to the objectives of utilization control. The same is true if budget allocations to an institution depend on occupancy levels, and if savings from more prudent management cannot be retained by the institution. In some situations, it may be possible to vary such factors under reasonably controlled conditions and to study the consequences.

The nature and extent of participation in monitoring and control by an external agency is probably a critical factor in effectiveness. Such an

external agency might be an insurance carrier, a government agency, or a PSRO. In a system such as that of the VA, the central office probably bears a parallel relationship to the local hospital. The relationship between the external agency and the intra-institutional quality monitoring apparatus can be structured in a variety of ways, and the consequences examined. For example, under the PSRO program, the hospital may be delegated, partially delegated or nondelegated with respect to quality and utilization control activities. The PSRO, or the analogous external agency, could obtain reports from the institution under its supervision, duplicate certain review activities, independently obtain data additional to those available to each institution, analyze and interpret data, engage in consultatory and educational activities, and so on. It may even be feasible to have hospitals take turns reviewing each other. The effects of each of these activities, and of the various ways in which each can be implemented, are legitimate and important objects of research.

Social-organizational characteristics and medical technology Based on more general work in organizational behavior, Scott et al. have hypothesized that a given organizational form may be more or less effective in influencing performance depending on the technological characteristics of the work being done.[72] Somewhat along the same line, though independent of the work of Scott et al., Wyszewianski has hypothesized that the effectiveness of a specific quality control procedure is influenced, among other things, by the degree to which it fits specified technological characteristics of the task whose performance it is designed to influence.[156] Further research along these lines may provide insights into the effectiveness of discrete quality control practices, while making a contribution to organizational theory.

Case studies of successful and unsuccessful programs. There is lively debate about the usefulness of case studies as a method for research. The contention that case studies are a valuable method for generating hypotheses is countered by the argument that it is impossible to understand anything about real life situations unless they are viewed and interpreted in preformed ways. Case studies cannot generate knowledge out of ignorance. However, given some general conceptual framework that suggests to the investigator where to look and what kinds of things to look for, case studies, I believe, can nudge the mind into new formulations that draw on what was formerly latent or only dimly perceived in the investigator's thought. Besides, the participants in the events that are under study have their own views and explanations of what these events mean. These are, in a way, fragments of theory and snippets of hypotheses; and in these bits and scraps the alert investigator may find the rough outlines of some new insight.

It seems reasonable to begin the investigation by examining sharply contrasting situations of brilliant success and abject failure. The factors that contribute to effectiveness should be more sharply differentiated by these large contrasts. But, further exploration through the examination of intermediate situations would have to follow. And all this must lead to more rigorous testing of specific hypotheses.

Cost-effectiveness and cost-benefit studies So far, in this section, I have dealt with documenting changes in behavior and in identifying the factors that seem to enhance or deter such changes. But, ultimately, a determination has to be made whether any given monitoring mechanism is worthwhile, and which among alternative mechanisms is to be preferred. This requires the comparison at least of costs and effects, and, preferably, of costs and benefits. For example, a committee of the Institute of Medicine has estimated that an extension of PSRO activities to cover all inpatient and ambulatory care would require a yearly expenditure of 1¼ billion dollars.[116] What do we get in return?

The returns to a program of monitoring quality and utilization are partly monetary savings due to reduction in "unnecessary" care. These savings have to be set aginst the expenditures that might result from care that is added in order to improve quality, but also due to redundant care masquerading as "quality." I have already discussed briefly the difficulties in measuring these savings and expenditures, and emphasized the need to clearly identify their social incidence. When that is done, the cost of establishing and operating the monitoring apparatus must be carefully determined so it can be set against the monetary balance of its consequences. But, no matter how careful and accurate the balance sheet of savings and expenditures is, a definitive judgment cannot be made unless it is possible to measure the impact on health. The measurement of health and the assigning of a monetary value to it has also been discussed in an earlier section. Seeing how difficult a task this is, we need not wonder that the effectiveness of so much that is done in medical care and its organization remains less than completely evaluated.

Epilogue

28 The task of reviewing the extent of current ignorance and of indicating ways of remedying it calls for an approach that the reader may find overly critical of what has been accomplished, and of what can yet be done, in the world of action. As I said at the beginning, it would be foolish to argue that all efforts to monitor quality must cease while we seek certainty about a near-perfect solution. On the contrary, we must continue to act based on what we now believe to be reasonable and feasible. But, we also need to find out whether we have been correct in acting as we have, and to learn how to do better in the future. I hope that this paper has made a small contribution to that continuing quest.

References

1. Brook, R H. and Allyson-Davies, A., *Quality Assurance and Cost Control in Ambulatory Care.* Santa Monica: The Rand Corporation, July 1977.
2. Tompkins, R.K.; Burner, D.C. and Cable, W.E., "An Analysis of Cost-Effectiveness of Pharyngitis Management and Acute Rheumatic Fever Prevention." *Annals of Internal Medicine* 86:481-492, April 1977.
3. Forsyth, R.A., "Selective Utilization of Clinical Diagnosis in Treatment of Pharyngitis." *The Journal of Family Practice* 2:173-177, June 1975.
4. Lusted, L.B., *Introduction to Medical Diagnosis.* Springfield, Illinois: Charles C. Thomas, 1968.
5. McNeil, B.J.; Keeler, E. and Adelstein, S.J., "Primer on Certain Elements of Medical Decision Making." *New England Journal of Medicine* 293: 211-215, July 31, 1975.
6. Raiffa, H., *Decision Analysis: Introductory Lectures on Choices Under Uncertainty.* Reading, Mass.: Addison-Wesley Publishing Company, 1968.
7. Klarman, H.E., "Application of Cost-Benefit Analysis to the Health Services and the Special Case of Technological Innovation." *International Journal of Health Services* 4:325-352, Spring 1974.
8. Bunker, J.P., Barnes, A.A. and Mosteller, F., Editors, *Costs, Risks and Benefits of Surgery.* New York: Oxford University Press, 1977.
9. Mushlin, A.I.; Appel, F.A. and Barr, D.M. "Quality Assurance in Primary Care: A Strategy Based on Outcome Assessment." *Journal of Community Health,* in press.
10. Jacobs, C.M.; Christoffel, T.H. and Dixon, N., *Measuring the Quality of Patient Care: The Rationale for Outcome Audit.* Cambridge, Mass.: Ballinger Publishing Company, 1976.
11. Weed, L.L., "Quality Control and the Medical Record." *Archives of Internal Medicine* 127:101-105, January 1971.
12. Williamson, J.W., "Evaluating Quality of Patient Care: A Strategy Relating Outcome and Process Assessment." *The Journal of the American Medical Association* 218:564-569, October 25, 1971.
13. Brook, R.H.; Davies-Avery, A.; Greenfield, S.; Harris, L.J.; Lelah, T.; Solomon, N.E. and Ware, J.E. Jr., "Assessing the Quality of Medical Care Using Outcome Measures: An Overview of the Method." *Medical Care* Supplement to Vol. 15, No. 9, September 1977. 165 pp.
14. Fanshel, S. and Bush, J.W., "A Health Status Index and Its Application to Health Services Outcome." *Operations Research* 18:1021-1060, November-December, 1970.
15. Hirsch, E.O., "Utilization Review as a Means of Continuing Education." *Medical Care* 12:358-362, April 1974.
16. McNeil, B.J.; Varady, P.D.; Burrows, B.A. and Adelstein, S.J., "Cost-Effectiveness Calculations in the Diagnosis and Treatment of Hypertensive Renovascular Disease." *New England Journal of Medicine* 293:216-221, July 31, 1975.
17. McNeil, B.J.; Hessel, S.J.; Branch, W.T.; Bjork, L. and Adelstein, S.J., "Measures of Clinical Efficacy. III. The Value of the Lung Scan in the Evaluation of Young Patients with Pleuritic Chest Pain." *Journal of Nuclear Medicine* 17:163-169, March 1976.
18. Neutra, R., "Indications for the Surgical Treatment of Suspected Acute Appendicitis: A Cost-Effectiveness Approach." In Bunker, J.P.; Barnes, B.A. and Mosteller, F., Editors, *Costs, Risks and Benefits of Surgery.* New York: Oxford University Press, 1977.
19. Schwartz, W.B., Gorry, G.A., Kassirer, J.P. and Essig, A., "Decision Analysis and Clinical Judgment." *The American Journal of Medicine* 55:459-472 (October 1973).
20. McNeil, B.J. and Adelstein, S.J., "The Value of Case Finding in Hypertensive Renovascu-

lar Disease." *New England Journal of Medicine* 293:221-226 (July 31, 1975).

21. Abt, C.C., "The Issue of Social Costs in Cost-Benefit Analysis of Surgery." In Bunker, J.P.; Barnes, B.A. and Mosteller, F., Editors, *Costs, Risks and Benefits of Surgery*. New York: Oxford University Press, 1977.

22. Ginsberg, A.S., *Decision Analysis in Clinical Patient Management with an Application to the Pleural-Effusion Syndrome*. Santa Monica, Calif.: The Rand Corporation, July 1971.

23. Meyers, A.; Brand, D.A.; Dove, H.G. and Dolan, T.F. Jr., *A Technique for Analyzing Clinical Data to Provide Patient Management Guidelines: A Study of Meningitis in Children*. New Haven, Conn.: Center for the Study of Health Services, Yale University, January 1975.

24. Sturdevant, R.A.L. and Stern, D., "Accuracy of Physicians' Predictions of Cholecystography Results." *Medical Care* 15:488-493, June 1977.

25. Piachaud, D. and Weddell, J.M., "The Economics of Treating Varicose Veins." *International Journal of Epidemiology* 1:287-294, Autumn, 1972.

26. Neuhauser, D. and Lewicki, A.M., "What Do We Gain From the Sixth Stool Guaiac?" *New England Journal of Medicine* 293:226-228, July 31, 1975.

27. Hardwick, D.F.; Vertinsky, P.; Barth, R.T.; Mitchell, V.F.; Bernstein, M. and Vertinsky, I., "Clinical Style and Motivation: A Study of Laboratory Test Use." *Medical Care* 13:397-408, May 1975.

28. Martin, S.P.; Donaldson, M.C.; London, C.D.; Peterson, O.L. and Colton, T., "Inputs into Coronary Care During 30 Years, A Cost Effectiveness Study." *Annals of Internal Medicine* 81:289-293, September 3, 1974.

29. Jackson, R.A. and Smith, M.C., "Relations Between Price and Quality in Community Pharmacy." *Medical Care* 12:32-39, January 1974.

30. Schroeder, S.A.; Schliftman, A. and Piemme, T.E., "Variation Among Physicians in Use of Laboratory Tests: Relation to Quality of Care." *Medical Care* 12:709-713, August 1974.

31. Rubenstein, L.; Mates, S. and Sidel, V.W., "Quality-of-Care Assessment by Process and Outcome Scoring: Use of Weighted Algorithmic Assessment Criteria for Evaluation of Emergency Room Care of Women with Symptoms of Urinary Tract Infection." *Annals of Internal Medicine* 86:617-625, May 1977.

32. Bloom, S.W. and Wilson, R.N., "Patient-Practitioner Relationships." In Freeman, H.E.; Levine, S. and Reeder, L.G., *Handbook of Medical Sociology*, Englewood Cliffs, N.J.: Prentice-Hall Inc., Second Edition, 1972.

33. Lebow, J.L., "Consumer Assessments of the Quality of Medical Care." *Medical Care* 12:328-337, April 1974.

34. Inui, T.S.; Yourtee, E.L. and Williamson, J.W., "Improved Outcomes in Hypertension After Physician Tutorials: A Controlled Trial." *Annals of Internal Medicine* 84:846-851, June 1976.

35. Sackett, D.L.; Gibson, E.S.; Taylor, D.W.; Haynes, R.B.; Hackett, B.C.; Roberts, R.S. and Johnson, A.L., "Randomised Clinical Trial of Strategies for Improving Medication Compliance in Primary Hypertension." *The Lancet* 2:1205-1207, May 31, 1975.

36. Smith, D.B. and Metzner, C.A., "Differential Perceptions of Health Care Quality in Prepaid Group Practice." *Medical Care* 8:264-275, July-August 1970.

37. Marram, B.D., "Patients' Evaluations of Nursing Performance." *Nursing Research* 22:153-157, March-April 1973.

38. DeGeyndt, W., "Five Approaches to Assessing the Quality of Care." *Hospital Administration* 15:21-42, Winter 1970.

39. Donabedian, A., "Some Basic Issues in Evaluating the Quality of Health Care." *Issues in Evaluation Research*. Kansas City, Mo.: American Nurses' Association, 1976.

40. Shortell, S.M., "Continuity in Medical Care: Conceptualization and Measurement. *Medical Care* 14:377-391, May 1975.

41. Bass, R.D. and Windle, C., "Continuity of Care: An Approach to Measurement." *American Journal of Psychiatry* 129:110-115, August 1972.

42. Pugh, T.F. and McMahon, B., "Measurement of Discontinuity in Psychiatric Inpatient Care." *Public Health Reports* 82:533-538, June 1967.

43. Mindlin, R.L. and Densen, P.M., "Medical Care of Urban Infants: Continuity of Care." *American Journal of Public Health* 59:1294-1301, August 1969.

44. Breslau, N. and Reeb, R.G., "Continuity of Care in a University-Based Practice." *Journal of Medical Education* 50:965-969, October 1975.

45. Breslau, N. and Haug, M.R., "Service Delivery Structure and Continuity of Care: A Case Study of Pediatric Practice in Process of

Reorganization." *Journal of Health and Social Behavior* 17:339-352, December 1976.

46. Starfield, B.H.; Simborg, D.W.,; Horn, S.D. and Yourtee, S.A., "Continuity and Coordination in Primary Care: Their Achievement and Utility." *Medical Care* 14:625-636, July 1976.

47. Bice, T.W. and Boxerman, S.B., "A Quantitative Measure of Continuity of Care." *Medical Care* 15:347-349, April 1977.

48. Shorr, G.I. and Nutting, P.A., "A Population-Based Assessment of the Continuity of Ambulatory Care." *Medical Care* 15:455-464, June 1977.

49. Becker, M.H.; Drachman, R.H. and Kirscht, J.P., "A Field Experiment To Evaluate Various Outcomes of Continuity of Physician Care." *American Journal of Public Health* 64:1062-1070, November 1974.

50. Childs, A.W. and Hunter, E.D., "Non-Medical Factors Influencing Use of Diagnostic X-Ray by Physicians." *Medical Care* 10:323-335, July-August, 1972.

51. Freeborn, D.K.; Baer, D.; Greenlick, M.R. and Bailey, J.W., "Determinants of Medical Care Utilization: Physicians' Use of Laboratory Services." *American Journal of Public Health* 62:846-853, June 1972.

52. Lyle, C.B. Jr.; Citron, D.S.; Sugg, W.C. and Williams, O.D., "Cost of Medical Care in a Practice of Internal Medicine: A Study in a Group of Seven Internists." *Annals of Internal Medicine* 81:1-6, July 1964.

53. Lyle, C.B. Jr.; Applegate, W.B.; Citron, D.S. and Williams, O.D., "Practice Habits in a Group of Eight Internists." *Annals of Internal Medicine* 84:594-601, May 1976.

54. Daniels, M. and Schroeder, S.A., "Variation Among Physicians in Use of Laboratory Tests. II. Relation to Clinical Productivity and Outcome of Care." *Medical Care* 15:482-487, June 1977.

55. Schroeder, S.; Kenders, K.; Cooper, J.K. and Piemme, T.E., "Use of Laboratory Tests and Pharmaceuticals: Variations Among Physicians and Effect of Cost Audit on Subsequent Use." *Journal of the American Medical Association* 225:969-973, August 20, 1973.

56. Fattu, N.C., "Experimental Studies of Problem Solving." *Journal of Medical Education* 39:212-225, February 1964.

57. Kleinmuntz, B., "The Processing of Clinical Information by Man and Machine." In Muntz, B.K., Editor, *Formal Representation of Human Judgment.* New York: John Wiley and Sons, Inc., 1968.

58. Leaper, D.J.; Gill, P.W.; Staniland, J.R.; Horrocks, J. and DeDombal, F.T., "Clinical Diagnostic Process: An Analysis." *British Medical Journal* 3:569-574, September 15, 1973.

59. Hare, R.L. and Barnoon, S., *Medical Care Appraisal and Quality Assurance in the Office Practice of Internal Medicine.* San Francisco: American Society of Internal Medicine, July 1973.

60. McDonald, C.J., "Protocol-Based Computer Reminders, The Quality of Care and the Non-Perfectability of Man." *New England Journal of Medicine* 95:1351-1355, December 9, 1976.

61. Brown, C.R. and Uhl, S.M., "Mandatory Continuing Education: Sense or Nonsense?" *Journal of the American Medical Association* 213:1660-1668, September 7, 1970.

62. Kessner, D.M.; Kalk, C.E. and Singer, J., "Assessing Health Quality—The Case for Tracers." *New England Journal of Medicine* 288:189-194, January 25, 1973.

63. Kessner, D.M.; Snow, C.K. and Singer, J., *Assessment of Medical Care For Children.* Washington, D.C.: Institute of Medicine, 1974.

64. Lyons, T.F. and Payne, B.C., "Interdiagnosis Relationships of Physician Performance Measures." *Medical Care* 12:369-374, April 1974.

65. Lyons, T.F. and Payne, B.C., "Interdiagnosis Relationships of Physician Performance Measures in Hospitals." *Medical Care* 15:475-481, June 1977.

66. Osborne, C.E., "Interdiagnosis Relationships of Physician Recording in Ambulatory Child Health Care." *Medical Care* 15:465-474, June 1977.

67. Payne, B.C.; Lyons, T.F.; Dwarshius, L.; Kolton, M. and Morris, W., *The Quality of Medical Care: Evaluation and Improvement.* Chicago: Hospital Research and Educational Trust, 1976.

68. Peterson, O.L.; Andrews, L.P.; Spain, R.S. and Greenburg, B.G., *"An Analytical Study of North Carolina General Practice, 1953-1954." Journal of Medical Education* Volume 31, No. 12, December 1956, Part 2.

69. Morehead, M.A.; Donaldson, R. et al., *A Study of the Quality of Hospital Care Secured by A Sample of Teamster Family Members in New York City.* New York: Columbia University School of Public Health and Administrative Medicine, 1964.

70. Brook, R.H. and Williams, K., "Evaluation of

the New Mexico Peer Review System, 1971 to 1973." *Medical Care* Vol. 14, No. 12, December 1976, Supplement. 122 pp.

71. Rhee, S., "Factors Determining the Quality of Physician Performance in Patient Care." *Medical Care* 14:733-750, September 1976.
72. Scott, W.R.; Forrest, W.H. Jr. and Brown, B.W., "Hospital Structure and Postoperative Mortality and Morbidity." In Shortell, S.M. and Brown, M., Editors, *Organizational Research in Hospitals*. Chicago: Blue Cross Association, 1976.
73. Brook, R.H. and Williams, K.N., "Quality of Health Care for the Disadvantaged." *Journal of Community Health* 1:132-156, Winter 1975.
74. Lyons, T.F. and Payne, B.C., "The Quality of Physicians' Health Care Performance: A Comparison Against Optimal Criteria for Treatment of the Elderly and Younger Adults in Community Hospitals." *Journal of the American Medical Association* 227:925-928, February 25, 1974.
75. Lyons, T.F. and Payne, B.C., "Quality of Physician' Office-Care Performance: Different for Elderly Than for Younger Adults?" *Journal of the American Medical Association*, 229:1621-1622, September 16, 1974.
76. Griner, P.F. and Liptzin, B., "Use of the Laboratory in a Teaching Hospital: Implications for Patient Care, Education, and Hospital Costs." *Annals of Internal Medicine* 75:157-163, August 1971.
77. Graham, J.B. and Paloucek, F.P., "Where Should Cancer of the Cervix Be Treated?" *American Journal of Obstetrics and Gynecology* 87:405-409, October 1963.
78. Schaffner, W.; Federspiel, C.F.; Fulton, M.L.; Gilbert, D.G. and Stevenson, L.B., "Maternal Mortality in Michigan: An Epidemiologic Analysis, 1950-1971." *American Journal of Public Health* 67:821-829, September 1977.
79. Hollingshead, A.B. and Redlich, F.C., *Social Class and Mental Illness: A Community Study*. New York: John Wiley & Sons, Inc., 1958.
80. Duff, R.S. and Hollingshead, A.B., *Sickness and Society*. New York: Harper and Row, 1968.
81. Rutstein, D.D.; Berenberg, W.; Chalmers, T.C.; Child, C.G. 3rd; Fishman, A.P. and Perrin, E.B., "Measuring the Quality of Medical Care: A Clinical Method." *New England Journal of Medicine* 294:582-588, March 11, 1976.
82. Donabedian, A., *Aspects of Medical Care Administration: Specifying Requirements for Health Care*. Cambridge, Mass.: Harvard University Press, 1973.
83. Patrick, D.L.; Bush, J.W. and Chen, M.M., "Toward an Operational Definition of Health." *Journal of Health and Social Behavior* 14:6-23, March 1973.
84. Patrick, D.L.; Bush, J.W. and Chen, M.M., "Methods for Measuring Levels of Well-being for a Health Status Index." *Health Services Research* 8:228-245, Fall 1973.
85. Torrance, G.W.; Thomas, W.H. and Sackett, D.L., "A Utility Maximization Model for Evaluation of Health Care Programs." *Health Services Research* 7:118-133, Summer 1972.
86. Gilson, B.S.; Gilson, J.S., Bergner M. et al., "The Sickness Impact Profile: Development of an Outcome Measure of Health Care." *American Journal of Public Health* 65:1304-1310, December 1975.
87. Pollard, W.E.; Bobbitt, R.A.; Bergner, M.; Martin, D.P. and Gilson, B.S., "The Sickness Impact Profile:Reliability of a Health Status Measure." *Medical Care* 14:146-155, February 1976.
88. Berg, R.L., Editor, *Health Status Indexes*. Chicago: Hospital Research and Educational Trust, 1973.
89. Elinson, J., Editor, "Sociomedical Health Indicators." *International Journal of Health Services* 6:337-538, 1976.
90. Several Authors, "Health Status Indexes—Work in Progress." *Health Services Research* 11:332-528, Winter, 1976.
91. Breslow, L., "A Quantitative Approach to the World Health Organization Definition of Health: Physical, Mental and Social Well-being." *International Journal of Epidemiology* 1:347-355, Winter 1972.
92. Codman, E.A., *A Study in Hospital Efficiency As Demonstrated by the Case Report of the First Five Years of a Private Hospital*. Boston: Thomas Todd Co., Circa 1916.
93. Hulka, B.S.; Zyzanski, S.J.; Cassel, J.C. and Thompson, S.J., "Scale for the Measurement of Attitudes toward Physicians and Primary Medical Care." *Medical Care* 8:429-435, September-October 1970.
94. Hulka, B.S.; Zyzanski, S.J.; Cassel, J.C. and Thompson, S.J., "Satisfaction with Medical Care in a Low Income Population." *Journal of Chronic Disease* 24:661-673, November 1971.
95. Zyzanski, S.J.; Hulka, B.S. and Cassel, J.C., "Scale for the Measurement of 'Satisfaction' with Medical Care: Modifications in Content, Format and Scoring." *Medical Care* 12:611-620, July 1974.

96. Schmidt, E.C.; Schall, D.W. and Morrison, C.C., "Computerized Problem-Oriented Medical Record For Ambulatory Practice." *Medical Care* 12:316-327, April 1974.

97. Grimm, R.H.; Shimoni, K; Harlan, W.R. and Estes, E.H. Jr., "Evaluation of Patient-Care Protocol Use by Various Providers." *New England Journal of Medicine* 292:507-511, March 6, 1975.

98. Lembcke, P.A., "Medical Auditing by Scientific Methods." *Journal of the American Medical Association* 162:646-655, October 13, 1956.

99. Wiener, S. and Nathanson, M., "Physical Examination: Frequently Observed Errors." *Journal of the American Medical Associaton* 236:852-855, August 16, 1976.

100. Turner, E.V.; Helper, M.M. and Kriska, S.D., "Predictors of Clinical Performance." *Journal of Medical Education* 47:959-965, December 1972.

101. Zuckerman, A.E.; Starfield, B.; Hochreiter, C. and Kovasznay, B., "Validating the Content of Pediatric Outpatient Medical Records by Means of Tape-Recording Doctor-Patient Encounters." *Pediatrics* 56:407-411, September 1975.

102. Steward, M.A. and Buck, S.W., "Physicians' Knowledge of and Response to Patients' Problems." *Medical Care* 15:578, July 1977.

103. Tufo, H.M.; Eddy, W.M.; Van Buren, H.C.; Bouchard, R.W.; Twitchell, J.C. and Bedard, L., "Audit in a Practice Group." Pages 29-41 in Walker, H.K.; Hurst, J.W. and Woody, M.F., Editors, *Applying the Problem-Oriented System*. New York: Medcom Press, 1973.

104. Simborg, D.W.; Starfield, B.H.; Horn, S.D. and Yourtee, S.A., "Information Factors Affecting Problem Follow-Up in Ambulatory Care." *Medical Care* 14:848-856, October 1976.

105. Grover, M. and Greenberg, T., "Quality of Care Given to First Time Birth Control Patients at a Free Clinic." *American Journal of Public Health* 66:986-987, October 1976.

106. Wassertheil-Smoller, S.; Bell, B. and Blaufox, M.D., "Computer-Aided Management of Hypertensive Patients." *Medical Care* 13:1044-1054, December, 1975.

107. Barnett, G.O.; Winickoff, R.; Dorsey, J.L. and Morgan, M., "The Role of Feedback in Quality Assurance—An Application of Computer-Based Ambulatory Medical Information System." In *Assessing Physician Performance in Ambulatory Care*. San Francisco: American Society of Internal Medicine, 1976.

108. Rosenberg, E.W., "What Kind of Criteria?" *Medical Care* 13:966-975, November 1975.

109. Greenfield, S.; Lewis, C.E.; Kaplan, S.H. and Davidson, M.B., "Peer Review by Criteria Mapping: Criteria for Diabetes Mellitus, The Use of Decision-Making in Chart Audit." *Annals of Internal Medicine* 83:761-770, December 1975.

110. Richardson, F.McD., "Methodological Development of a System of Medical Audit." *Medical Care* 10:451-462, November-December 1972.

111. Hoekelman, R.A. and Peters, E.N., "A Health Supervision Index to Measure Standards of Child Care." *Health Services Reports* 87:537-544, June-July 1972.

112. Lyons, T.F. and Payne, B.C., "The Use of Item Importance Weights in Assessing Physician Performance with Predetermined Criteria Indices." *Medical Care* 13:432-439, May 1975.

113. Hopkins, C.E.; Hetherington, R.W. and Parsons, E.M., "Quality of Medical Care: A Factor Analysis Approach Using Medical Records." *Health Services Research* 10:199-208, Summer 1975.

114. Novick, N.R.; Dickinson, K; Asnes, R.; Lan, M. S-P. and Lowenstein, R., "Assessment of Ambulatory Care: Application of Tracer Methodology." *Medical Care* 14:1-12, January 1976.

115. Donabedian, A. and Wyszewianski, L., "The Numerology of Utilization Control Revisited: When to Recertify." *Inquiry* 14:96-102, March 1977.

116. Institute of Medicine, *Assessing Quality in Health Care: An Evaluation*. Washington, D.C.: National Academy of Sciences, November 1976.

117. Averill, R.F. and McMahon, L.F., "A Cost Benefit Analysis of Continued Stay Certification." *Medical Care* 15:158-173, February 1977.

118. Wolfe, H., "A Computerized Screening Device for Selecting Cases for Utilization Review." *Medical Care* 5:44-51, January-February 1967.

119. Riedel, D.C.; Fetter, R.B.; Mills, R.E. and Pallett, P.J., *Basic Utilization Review Program (BURP)*. New Haven, Conn.: Health Services Research Program, Yale University, 1973.

120. Rubin, L., *Comprehensive Quality Assurance System: The Kaiser-Permanente Approach*. Alexandria, Va.: American Group Practice Association, 1975.

121. Glass, R.I.; Mulvhill, M.N.; Smith, H. Jr.; Peto, R.; Bucheister, D. and Stoll, B.J., "The

4 Score: An Index for Predicting A Patients' Non-Medical Hospital Days." *American Journal of Public Health* 67:751-755, August 1977.

122. Brauer, L.D., "Selection of Cases for Individual Review." In Riedel, D.C.; Tischler, G.L. and Myers, J.K., *Patient Care Evaluation in Mental Health Programs.* Cambridge, Mass.: Ballinger Publishing Company, 1974.

123. Donabedian, A., *A Guide to Medical Care Administration, Volume II: Medical Care Appraisal–Quality and Utilization.* New York (now Washington): American Public Health Association, 1969. 176 pp. (plus Annotated Selected Bibliography by A.J. Anderson).

34

124. Brook, R.H. and Appel, F.A., "Quality-of-Care Assessment: Choosing a Method for Peer Review." *New England Journal of Medicine* 288:1323-1329, June 21, 1973.

125. Brook, R.H., *Quality of Care Assessment: A Comparison of Five Methods of Peer Review.* Washington, D.C.: U.S. Department of Health, Education and Welfare, Bureau of Health Services Research and Evaluation, July 1973.

126. New York Academy of Medicine, Committee on Public Health Relations, *Maternal Mortality in New York City: A Study of All Puerperal Deaths, 1930-32.* The Commonwealth Fund, New York: Oxford University Press, 1933.

127. New York Academy of Medicine, Committee on Public Health Relations, *Perinatal Mortality in New York City: Responsible Factors.* Cambridge, Mass.: Harvard University Press, for the Commonwealth Fund, 1955.

128. Oakley, J.R.; McWeeny, P.M.; Hayes-Allen, M. and Emery, J.L., "Possibly Avoidable Deaths in Hospital in the Age-Group One Week to Two Years." *The Lancet* 1:770-772, April 10, 1976.

129. Gonnella, J.S. and Goran, M.J., "Quality of Patient Care—A Measurement of Change: The Staging Concept." *Medical Care* 13:467-473, June 1975.

130. Gonnella, J.S.; Louis, D.Z. and McCord, J.J., "The Staging Concept—An Approach to the Assessment of Outcome of Ambulatory Care." *Medical Care* 14:13-21, January 1976.

131. Brook, R.H. and Stevenson, R.L., "Effectiveness of Patient Care in an Emergency Room." *New England Journal of Medicine* 283:904-907, October 22, 1970.

132. Starfield, B. and Scheff, D., "Effectiveness of Pediatric Care: The Relationship Between Process and Outcome." *Pediatrics* 49:547-552, April 1972.

133. McCarthy, E.G. and Widmer, G.W., "Effects of Screening by Consultants on Recommended Elective Surgical Procedures." *New England Journal of Medicine* 291:1331-1335, December 19, 1974.

134. McCarthy, E.G., "Mandatory and Voluntary Second Opinion Programs in the Greater New York Area with National Implications." The Dorothy Eisenberg Lecture, Department of Surgery, Harvard Medical School, June 12, 1976.

135. Freeborn, D.K. and Greenlick, M.R., "Evaluation of the Performance of Ambulatory Care Systems: Research Requirements and Opportunities." *Medical Care* 11:68-75, March-April 1973.

136. Christoffel, T. and Lowenthal, H., "Evaluating the Quality of Ambulatory Health Care: A Review of Emerging Methodologies." *Medical Care* 15:877-897, November 1977.

137. American Society of Internal Medicine, *Assessing Physician Performance in Ambulatory Care.* San Francisco: The Society, 1976.

138. American Nurses' Association, *Issues in Evaluation Research.* Kansas City, Mo.: The Association, 1976.

139. Several Authors, "Symposium on the Quality of Pharmaceutical Services." *Drugs in Health Care* 2:3-66, Winter 1975.

140. Knapp, D.A. and Smith, M.C., *Evaluating Pharmacists and Their Activities.* Washington, D.C.: American Society of Hospital Pharmacists, Research and Education Foundation, 1973.

141. Several papers in the *Journal of the American Dental Association* 89: 842-871, October 1974.

142. Riedel, D.C.; Tischler, G.L. and Myers, J.K., Editors, *Patient Care Evaluation in Mental Health Programs.* Cambridge, Mass.: Ballinger Publishing Company, 1974.

143. Ehrlich, J.; Morehead, M.A. and Trussell, R.E., *The Quantity, Quality and Costs of Medical and Hospital Care Secured by a Sample of Teamster Families in the New York Area.* New York: Columbia University, School of Public Health and Administrative Medicine, 1962.

144. Shenkin, B.N. and Warner, D.C., "Giving the Patient His Medical Record: A Proposal to Improve the System." *New England Journal of Medicine* 289:688-692, September 27, 1973.

145. Stevens, P.D.; Stagg, R. and Mackay, I.R., "What Happens When Hospitalized Patients See Their Own Records." *Annals of Internal Medicine* 86:474-477, April 1977.

146. Bouchard, R.E.; Tufo, H.M.; Van Buren, H.C.; Eddy, W.M.; Twitchell, J.C. and Bedard, L., "The Patient and His Problem-

Oriented Record." In Walker, H.K.; Hurst, J.W. and Woody, M.F., Editors, *Applying the Problem-Oriented System.* New York: Mecom Press, 1973.

147. Burger, C.S., "Audit in a Problem-Oriented Practice." In Walker, H.K.; Hurst, J.W. and Woody, M.F., Editors, *Applying the Problem-Oriented System.* New York: Medcom Press, 1973.

148. Goldblatt, P.B.; Henisz, J.E. and Tischler, G.L, "Utilization Review Within an Institutional Context." In Riedel, D.E.; Tischler, G.L. and Myers, J.K., Editors, *Patient Care Evaluation in Mental Health Programs.* Cambridge, Mass.: Ballinger Publishing Company, 1974.
pany, 1974. 292 pp.

149. Donabedian, A., "Promising Quality Through Evaluating the Process of Patient Care." *Medical Care* 6:181-202, May-June 1968.

150. Freidson, E., "Speculations on the Social Psychology of Local PSRO Operations." In *Proceedings: Conference on Professional Self-Regulation, June, 1975.* DHEW Publication No. (HRA) 77-621, Washington, D.C.

151. Morehead, M.A., "P.S.R.O.—Problems and Possibilities." *Man and Medicine* 1:113-132, Winter 1976.

152. Bellin, L.E., "PSRO—Quality Control? Or Gimmickry?" *Medical Care* 12:1012-1018, December 1974.

153. Brian, E.W., "Government Control of Hospital Utilization: A California Experience." *New England Journal of Medicine* 286:1340-1344, June 22, 1972.

154. Glendenning, M.K.; Wolfe, H.; Shuman, L.J. and Huber, G.A., "The Effect of a Target Date Based Utilization Review Program on Length of Stay." *Medical Care* 14:751-764, September 1976.

155. Lave, J.R. and Leinhardt, S., "An Evaluation of a Hospital Stay Regulatory Mechanism." *American Journal of Public Health* 66:959-967, October 1976.

156. Wyszewianski, L., work in progress, Department of Medical Care Organization, The University of Michigan, Ann Arbor.

National Center for Health Services Research publications of interest to the health community are available on request to NCHSR, Office of Scientific and Technical Information, 3700 East-West Highway, Room 7-44, Hyattsville, MD 20782 (telephone: 301/436-8970). Mail requests will be facilitated by enclosure of a self-addressed self-adhesive mailing label. These publications also are available for sale through the National Technical Information Service (NTIS), Springfield, VA 22161 (telephone: 703/557-4650). PB and HRP numbers in parentheses are NTIS order numbers. Publications which are out of stock in NCHSR are indicated as available only from NTIS. Prices may be obtained from the NTIS order desk on request.

Research Digests

The *Research Digest Series* provides overviews of significant research supported by NCHSR. The series describes either ongoing or completed projects directed toward high priority health services problems. Issues are prepared by the principal investigators performing the research, in collaboration with NCHSR staff. Digests are intended for an interdisciplinary audience of health services planners, administrators, legislators, and others who make decisions on research applications.

(HRA) 76-3144 Evaluation of a Medical Information System in a Community Hospital (PB 264 353)

(HRA) 76-3145 Computer-Stored Ambulatory Record (COSTAR) (PB 268 342)

(HRA)77-3160 Program Analysis of Physician Extender Algorithm Projects (PB 264 610, available NTIS only)

(HRA) 77-3161 Changes in the Costs of Treatment of Selected Illnesses, 1951-1964-1971 (HRP 0014598)

(HRA) 77-3163 Impact of State Certificate-of-Need Laws on Health Care Costs and Utilization (PB 264 352)

(HRA) 77-3164 An Evaluation of Physician Assistants In Diagnostic Radiology (PB 266 507, available NTIS only)

(HRA) 77-3166 Foreign Medical Graduates: A Comparative Study of State Licensure Policies (PB 265 233)

(HRA) 77-3171 Analysis of Physician Price and Output Decisions (PB 273 312)

(HRA) 77-3173 Nurse Practitioner and Physician Assistant Training and Deployment (PB 271 000) available NTIS only

(HRA) 77-3177 Automation of the Problem-Oriented Medical Record (PB 266 881)

Research Summaries

The *Research Summary Series* provides rapid access to significant results of NCHSR-supported research projects. The series presents executive summaries prepared by the investigators at the completion of the project. Specific findings are highlighed in a more concise form than in the final report. The *Research Summary Series* is intended for health services administrators, planners, and other research users who require recent findings relevant to immediate programs in health services.

(HRA) 77-3162 Recent Studies in Health Services Research, Vol. 1 (July 1974 through December 1976) (PB 266 460)

(HRA) 77-3183 Recent Studies in Health Services Research, Vol. II (CY 1976) (PB 279 198)

(HRA) 77-3176 Quality of Medical Care Assessment Using Outcome Measures (PB 272 455)

(HRA) 78-3193 Optimal Electrocardiography (PB 281 558)

Policy Research

The *Policy Research Series* describes findings from the research program that have major significance for policy issues of the moment. These papers are prepared by members of the staff of NCHSR or by independent investigators. The series is intended specifically to inform those in the public and private sectors who must consider, design, and implement policies affecting the delivery of health services.

(HRA) 77-3182 Controlling the Cost of Health Care (PB 266 885)

Research Reports

The *Research Report Series* provides significant research reports in their entirety upon the completion of the project. Research Reports are developed by the principal investigators who conducted the research, and are directed to selected users of health services research as part of a continuing NCHSR effort to expedite the dissemination of new knowledge resulting from its project support.

(HRA) 76-3143 Computer Based Patient Monitoring Systems (PB 266 508)

(HRA) 77-3152 How Lawyers Handle Medical Malpractice Cases (HRP 0014313)

(HRA) 77-3159 An Analysis of the Southern California Arbitration Project, January 1966 through June 1975 (HRP 0012456)

(HRA) 77-3165 Statutory Provisions for Binding Arbitration of Medical Malpractice Cases (PB 264 409 available NTIS only)

(HRA) 77-3184 1960 and 1970 Spanish Heritage Population of the Southwest by County (PB 280 656)

(HRA) 77-3188 Demonstration and Evaluation of a Total Hospital Information System (PB 271 079)

(HRA) 77-3189 Drug Coverage under National Health Insurance: The Policy Options (PB 272 074)

(PHS) 78-3204 Experiments in Interviewing Techniques: Field Experiments in Health Reporting (PB 276 080)

Research Management

The *Research Management Series* describes programmatic rather than technical aspects of the NCHSR research effort. Information is presented on the NCHSR goals, research objectives, and priorities; in addition, this series contains lists of grants and contracts, and administrative information on funding. Publications in this series are intended to bring basic information on NCHSR and its programs to research planners, administrators, and others who are involved with the allocation of research resources.

(HRA) 76-3136 The Program in Health Services Research (Revised 9/76)

(HRA) 77-3158 Summary of Grants and Contracts, Active June 30, 1976

(HRA) 77-3167 Emergency Medical Services Systems Research Projects (Active as of June 30, 1976) (PB 264 407, available NTIS only)

(HRA) 77-3179 Research on the Priority Issues of the National Center of Health Services Research, Grants and Contracts Active on June 30, 1976

(HRA) 77-3194 Emergency Medical Services Systems Research Projects, 1977 (PB 273 893)

(HRA) 78-3202 NCHSR Research Bibliography (July 1, 1976 to June 30, 1977) (PB 273 997)

Research Proceedings

The *Research Proceedings Series* extends the availability of new research announced at key conferences, symposia and seminars sponsored or supported by NCHSR. In addition to papers presented, publications in this series includes discussions and responses whenever possible. The series is intended to help meet the information needs of health services providers and others who require direct access to concepts and ideas evolving from the exchange of research results.

(HRA) 76-3138 Women and Their Health: Research Implications for a New Era (PB 264-359, available NTIS only)

(HRA) 76-3150 Intermountain Medical Malpractice (PB 268 344, available NTIS only)

(HRA) 77-3154 Advances in Health Survey Research Methods (PB 262 230)

(HRA) 77-3181 NCHSR Research Conference Report on Consumer Self-Care in Health (PB 273 811)

(HRA) 77-3186 International Conference on Drug and Pharmaceutical Services Reimbursement (PB 271 386)

(PHS) 78-3195 Emergency Medical Services: Research Methodology (PB 279 096)

Program Solicitations

(HRA) 77-3196 Conference Grant Information

(HRA) 77-3200 Grants for Dissertation Research Support

(PHS) 78-3206 Grants for Cost Containment Research for Health Planning

BIBLIOGRAPHIC DATA SHEET	1. Report No. NCHSR 78-145	2.	3. Recipient's Accession No.
4. Title and Subtitle NEEDED RESEARCH IN THE ASSESSMENT AND MONITORING OF THE QUALITY OF MEDICAL CARE; NCHSR Research Report Series			5. Report Date July 1978 6.
7. Author(s) Avedis Donabedian			8. Performing Organization Rept. No.
9. Performing Organization Name and Address The University of Michigan School of Public Health Department of Medical Care Organization Ann Arbor, MI 48104 AC 313/764-5433			10. Project/Task/Work Unit No. --- 11. Contract/Grant No. HS 02081
12. Sponsoring Organization Name and Address DHEW, PHS, OASH, National Center for Health Services Research 3700 East-West Highway, Room 7-44 (STIB) Hyattsville, MD 20782 AC 301/436-8970			13. Type of Report & Period Covered Interim Rpt. 5/1/76 - 4/30/79 14.

15. Supplementary Notes
DHEW Pub. No. (PHS) 78-3219. The final report will not be completed until approx. June 1979. This is an interim report.

16. Abstracts
The purpose of this report is to review, evaluate critically, and synthesize the literature on quality assessment and assurance, including the appropriateness of use of service, in order to arrive at a cogent, documented, and authoritative assessment of the state-of-the-art. In addition to addressing quality assessment as a research tool and quality assurance as an administrative tool, an attempt is made to provide an understanding of the epidemiology of quality as a prerequisite to the design of medical care programs and systems. Major components of quality which are discussed include: (1) definitions, (2) quality assessment and program evaluation, (3) relationship of quality and quantity, (4) relationship of quality and cost, (5) strategies of care, (6) structure, process, and outcome, (7) monitoring versus research, and (8) the uses of outcomes. Recommendations for further research in the assessment and monitoring of the quality of medical care are presented.

~~17. Key Words and Document Analysis. 17a. Descriptors~~

NCHSR publication of research findings does not necessarily represent approval or official endorsement by the National Center for Health Services Research or the U.S. Department of Health, Education, and Welfare.

Diane N. Funk: NCHSR P.O.; AC 301/436-8941.

17b. Identifiers/Open-Ended Terms
Health services research
Quality of medical care
Quality assessment
Quality assurance
Monitoring the quality of care

17c. COSATI Field/Group

18. Availability Statement	19. Security Class (This Report)	21. No. of Pages
Releasable to the public. Available from National Technical Information Service, Springfield, VA 22161 AC 703/557-4650	UNCLASSIFIED	Est.: 50
	20. Security Class (This Page) UNCLASSIFIED	22. Price

FORM NTIS-35 (REV. 10-73) ENDORSED BY ANSI AND UNESCO. THIS FORM MAY BE REPRODUCED USCOMM-DC 8265-P74